The Dissertation Committee for Lawrence Van Fulton

certifies that this is the approved version of the following dissertation:

Performance of Army Medical

Department Health Delivery Components,

2001-2003: A Multi-Model Approach

Committee:

Leon S. Lasdon, Supervisor

William W. Cooper, Co-Supervisor

Patrick L. Brockett

Reuben R. McDaniel, Jr.

Barbara Wojcik

Performance of Army Medical

Department Health Delivery Components,

2001-2003: A Multi-Model Approach

by

Lawrence Van Fulton, BS; MHA; MS; MMAS; MSStat

Dissertation

Presented to the Faculty of the Graduate School

of The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

May 2005

Acknowledgements

The efforts of my family and the committee members underwrote the successful completion of this dissertation. I am forever indebted. Thank you.

Performance of Army Medical

Department Health Delivery Components,

2001-2003: A Multi-Model Approach

Publication No.________

Lawrence Van Fulton, PhD

University of Texas at Austin, 2005

Supervisors: Leon S. Lasdon and William W. Cooper

Efficiency analyses of 24 military hospitals for the years 2001-2003 were performed using Data Envelopment Analysis (DEA), Stochastic Frontier Analysis (SFA), and Corrected Ordinary Least Squares (COLS) methodologies. The facilities under analysis represented $4 billion of annual Army Medical Department expenditures. The results generally identified the same high and low performance outliers. Output and input slacks and referent sets were analyzed to determine if patterns of performance existed.

Hospital cost models were then analyzed. Investigation of the optimal Box-Cox transformations for all variables resulted in the selection a simple loglinear cost function. Hospital cost was modeled as a function of workload, population, a quality and prevention proxy, an access proxy, efficiency scores (and interactions), medical center status, and the interaction between medical center status and workload.

Both cross-sectional and panel series studies were conducted. Traditional, ridge, and robust regression methods were applied to the models and compared with the Stochastic Frontier Analysis estimates of hospital cost. Estimation techniques included least squares, Markov Chain Monte Carlo (MCMC) simulation, and Maximum Likelihood Estimation. The results of the comparison indicated that linear models with DEA efficiency variables provided better estimates than SFA models. The best longitudinal model was unbiased with small variance and exhibited an extremely strong linear relationship ($R^2=.98$). The models provided evidence to support the following relationship: *Hospital Cost = f(Workload, Efficiency, Quality, MEDCEN, MEDCEN*Workload, Time*).

Using the models that demonstrated the smallest empirical variance and bias, parameter estimates generated at time t were then used to forecast cost at time $t+1$ under either the assumption of time invariant efficiency (for cross-sectional forecasts) or by employing a moving average efficiency score. These forecasts were then evaluated for efficacy.

Next, a method for adjusting the funding of facilities by using management directed efficiency minimums and estimation error was proffered. The recommendations associated with this method were compared against the slack variables from DEA analysis. Additional analysis of military hospital and network efficiency were also provided as the basis for future research.

Table of Contents

List of Tables

List of Figures

Chapter 1. Introduction

Overview of Research

Purpose

This dissertation analyzes performance of the Army Medical Department (AMEDD) military facilities and associated networks from 2001-2003. The objectives of this research are to evaluate cost, quality and prevention, access, training, readiness, and workload performance for AMEDD system components (hospitals, medical centers, and supporting network facilities) in order to make recommendations for possible resource shifts within and across military medical facilities that have a total annual budget of approximately $4 billion. The current method of budget allocation is primarily historically based plus an adjustment for inflation.

Population of Interest

The primary population of interest includes the 17 Army Community Hospitals and seven Army Medical Centers in the United States; these facilities are observed over the time period 2001-2003 (n=72). Separate analyses are conducted on the Medical Centers (n=21) and the hospitals (n=51).[1]

Approach

Up until the *2002 Evaluation of TRICARE Performance Report to Congress* (www.tricare.osd.mil), performance reports associated with the AMEDD focused on "before and after" comparisons. The previous evaluations compared estimates of

---

[1] Appendix 8 provides an initial analysis of the civilian healthcare provision to eligible beneficiaries in geographic areas surrounding each military facility (n = 144).

performance results that would have been obtained prior to the implementation of the TRICARE (the Department of Defense healthcare system) with results actually observed under the current system. The new evaluative approach, however, focuses on trend analysis post-implementation of TRICARE. Similarly, this dissertation focuses on current system performance under TRICARE.

Many of the most recent studies regarding health care performance in the military do not consider less tangible outputs of production such as quality / prevention, training, and readiness (e.g., physician deployments). Each of these components is a valid organizational output, which should be considered along with the typical components of production such as weighted workload. Further, the "readiness" component of the AMEDD mission is unique and not shared by the civilian network. This dissertation takes into consideration these outputs and provides some insight into the secondary and tertiary effects that readiness has on other outputs of production.

Several models and estimation techniques are used in order to avoid methodological bias. The models and the results are compared for consistency and are used to determine if any implications regarding resource re-allocation or adjustments might be made.

Motivation, Significance, and Applicability

The motivation for this dissertation is the requirement to provide performance-based, data-driven decision support for evaluating and budgeting AMEDD facilities. The United States Army Surgeon General and the Center for AMEDD Strategic Studies

(CASS) seek to optimize the allocation of resources based on system inputs and outputs. This dissertation supports their efforts and the Army Surgeon General's balanced scorecard initiative, which specifies and measures achievement of organizational goals.

Structure

The content of this dissertation is structured as follows. This chapter provides the necessary background information regarding the TRICARE system and the Army Medical Department. The literature review (Chapter 2) examines previous studies regarding military health systems and the associated methodologies. Chapter 3 develops the separate methodologies employed in this research, while Chapter 4 provides the data sources, variable transformations, and descriptive statistics. Chapter 5 provides the results of the efficiency analysis of military facilities, while Chapter 6 compares and contrasts different cost models. The final chapter is an executive overview of this work.

For those unfamiliar with the military health care system, an overview of both the Department of Defense and Army Medical Department is required. A cursory description of these systems is provided in the following section.

The TRICARE Military Health System

The Army Medical Department (AMEDD) provides care to military service members, their dependents, retirees, survivors and other authorized personnel underneath the umbrella of the Department of Defense TRICARE program. The TRICARE program is a congressionally mandated collection of health plans managed by geographic region. In each geographic region, there is a single Lead Agent, who is responsible for coordinating care in the military facilities and throughout the network. The regions include both military medical facilities and civilian network components. Large contracts for the provision of health care outside of the fixed facility medical entities are issued based upon these regions. Currently, the contractors are TriWest Healthcare Alliance, Humana Military, and Health Net Federal Services. The geographic regions recently realigned as depicted in Figure 1.

Figure 1. TRICARE Geographic Regions.
The TRICARE geographic regions realigned as the contracts were consolidated. The pictures derive from www.TRICARE.osd.mil, the TRICARE official website, and represent the changes from June 1, 2004 to November 1, 2004.

Each authorized beneficiary is provided with multiple options for receiving health care. The three primary options include TRICARE Prime, TRICARE Extra, and TRICARE Standard. These three options are the focus of this study and are discussed below. There are other programs including TRICARE for Life (a Medicare wrap-around program), TRICARE Prime Remote (for those who work and live in remote areas), and TRICARE Plus (Military Treatment Facility primary care services only); however, these programs are not included in this research.

TRICARE Prime. TRICARE Prime is a military health care plan in which the authorized beneficiary enrolls to receive health care either within a Military Treatment Facility or in a participating civilian network component. The plan is similar to a Health Maintenance Organization (HMO) plan in that the enrollee agrees to receive their care from a specific primary care team or provider, and the enrollee must first seek care from this provider for all non-urgent needs. A Health Care Finder facilitates referrals to the civilian network specialty care.

All Active Duty military personnel are required to enroll in TRICARE Prime, and all other authorized beneficiaries other than the Medicare eligible may enroll. There are no restrictions regarding pre-existing conditions.

TRICARE Prime enrollees are guaranteed access within time limits. Drive time to the primary care site should not exceed 30 minutes. Waiting times for acute care may not exceed one day, while waiting time for routine care should not exceed one week. Specialty care is to be available within a one-hour drive with a maximum wait of four weeks for an appointment. Emergency care is to be available at all times.

While TRICARE Prime is free to military personnel, the cost to retirees is $230 per retiree or family member with a maximum cost of $460 per annum. The co-pay for any visit is $12 unless care is provided inside the military facility. In this case, there is no co-payment required.

TRICARE Extra. TRICARE Extra is a traditional Preferred Provider Organization (PPO) plan that requires no enrollment. Beneficiaries select any provider that has agreed to accept the Extra reimbursement rates. TRICARE will then pay for 80% of the Maximum Allowable Charge (MAC) for Active Duty (AD) and their families and 75% of the MAC for retirees and their families. The patient is responsible for the difference. The provider agrees to accept the discounted MAC reimbursement rate so that there is no additional charge to the patient. The annual deductible is $150 per individual or $300 per family. Since this plan does not require enrollment, the patient may use any physician within who accepts either TRICARE Extra or TRICARE Standard.

TRICARE Standard. TRICARE Standard is the program formerly referred to as CHAMPUS. In this program, patients are responsible for 25% and 30% of the MAC for AD-related beneficiaries and retiree-related beneficiaries respectively. Further, physicians may charge more than the MAC, so patients may be responsible for the difference between MAC and the provider charges (up to 15% balance billing). The annual deductible is $150 per individual or $300 per family. Under TRICARE Standard, AD family members must pay $11.90 per day for inpatient hospital stays, while retirees and their family members pay $414 per day or 25% of the cost for any separately billed physician and professional fees, whichever is less. In addition, the program does not cover all health care; it only covers medically necessary care.

Although the TRICARE system incorporates Army, Air Force, and Navy medical facilities, the interest of this research is the Army facilities and the associated network entities. With the basic description of TRICARE in place, the discussion of the

Army Medical Department structure is necessary. More information about TRICARE is available at www.TRICARE.osd.mil.

The Army Medical Department Health care System

The Army Medical Department is a multi-billion dollar component of the United States Army, and provides both peacetime and wartime health care for authorized beneficiaries. The primary structure of the organization is depicted in Figures 2 and 3 (see www.armymedicine.army.mil for more structural information). This research is most interested in the Regional Medical Command (RMC) subordinate entities, which are the command elements for the Army Medical Centers (AMCs) and Army Community Hospitals (ACHs) (the focus of this research).

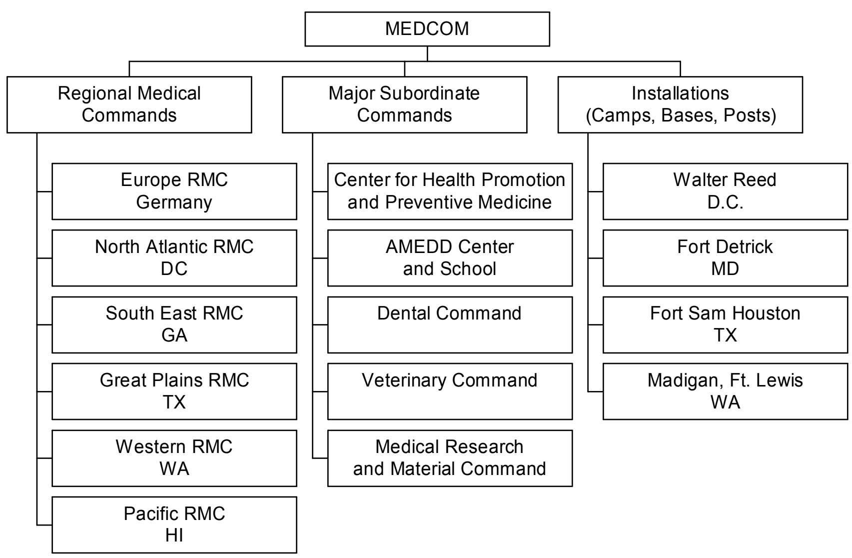


Figure 2. Medical Command Structure.

The structure of the Army Medical Command (MEDCOM) includes subordinate Regional Medical Commands responsible for the delivery of health care.

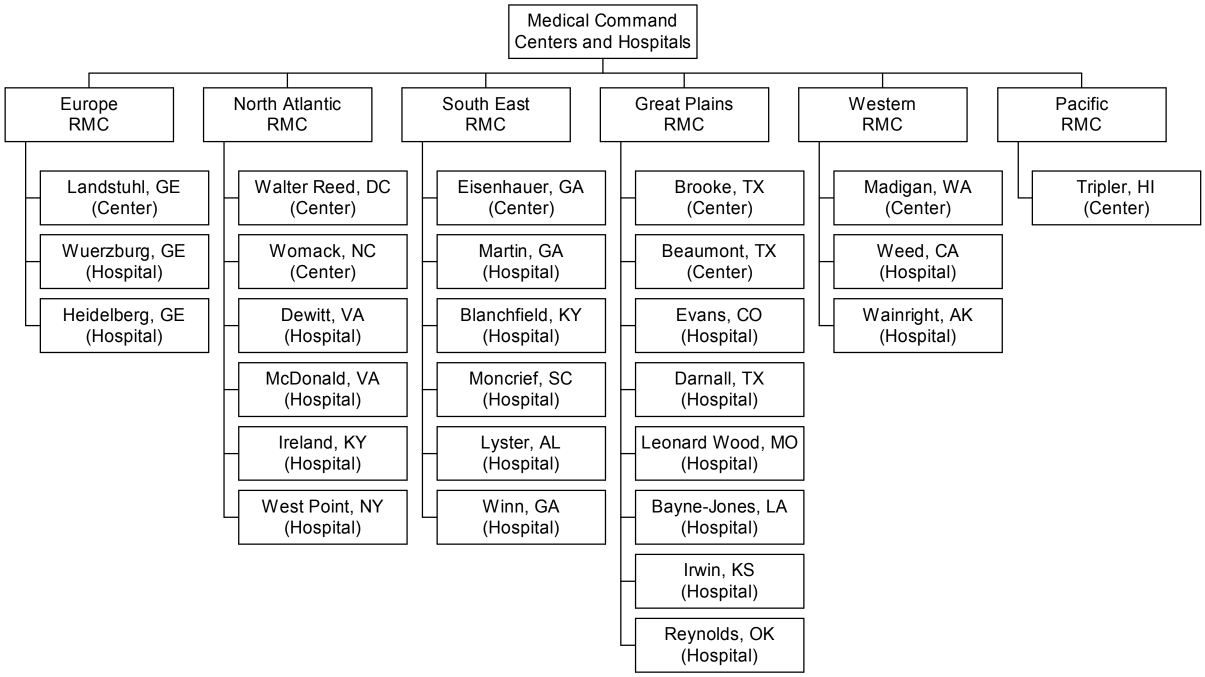


Figure 3. Hospitals in the AMEDD.

.

A complexity associated with the AMEDD structure is that the RMCs overlap multiple TRICARE lead agent regions. This structural peculiarity is a potential source of inefficiency in that a single RMC (such as Great Plains RMC) expends additional effort (if not resources) in assisting facilities, which are subject to different TRICARE regional contracts.

The size and the scope of the AMEDD's daily health care delivery mission for April 2002 through March 2003 are summarized in Table 1. These numbers serve to illustrate that the AMEDD is a large, multi-component health care entity.

Table 1. AMEDD Workload.

From April 2002 through March 2003, the AMEDD produced the following daily workload (http://www.armymedicine.army.mil). Food inspection is a responsibility of the Army veterinarians. Annual numbers are based on a 365-day calendar year.

	One Day	Year
patient beds occupied	1,295	472,675
clinic visits	37,217	13,584,205
dental visits	6,400	2,336,000
patients admitted	359	131,035
immunizations	5,462	1,993,630
births	63	22,995
laboratory procedures	49,226	17,967,490
veterinary outpatient visits	1,507	550,055
pharmacy procedures	69,524	25,376,260
millions worth of food inspected	$ 15.00	$ 5,475.00

This basic description of the AMEDD organization is sufficient for the study. Additional information regarding the structure and components of the Army Medical Department is available at www.armymedicine.army.mil.

Synopsis

The Army component of the TRICARE system includes those services provided by either the AMEDD Medical Treatment Facilities or the supporting TRICARE network system as provided by one of the three regional contractors. The system is geographically specific for both the network and the MTF components; however, the controlling agencies for the military components and the civilian components do not overlap. It is clear that the inherent complexity in such a system may be a source of inefficiency and waste. This inefficiency has been studied using several different methodologies in the recent past, and a review of both the studies and the methodologies will provide the basis for model development.

Chapter 2: Literature Review

In this chapter, a review of selected efficiency and effectiveness studies is provided. Specifically, several Data Envelopment Analysis (DEA) publications, Least Squares (LS) models, Stochastic Frontier Analysis (SFA), and Bayesian studies are discussed in order to provide the foundation for the development of reasonable efficiency and cost models.

Cooper, Seiford, and Tone (1999) suggest that effectiveness is the ability to state and achieve desired goals, while efficiency relates to the benefits realized and the resources used. The difference is somewhat subtle but very important. An organization might be efficient in that it is capable of providing many services with few resources; however, it may be ineffective, as the quality of those services might not achieve the desired goals. Conversely, a hospital may provide the best quality care and be categorized as effective, while consuming exceedingly large amounts of resources (inefficient). In this analysis, efficiency is defined as optimization of the relationship between resources and benefits. Effectiveness is indirectly taken into account by the choices of the inputs and outputs seen in the efficiency evaluation but is not otherwise addressed explicitly.

Multiple methods exist for evaluating cost efficiencies in military health care facilities. Data Envelopment Analysis (DEA) is often used to provide efficiency estimates for systems with multiple inputs and outputs. Stochastic Frontier Analysis (SFA) is used to provide both efficiency estimates and cost estimates in (normally) single output models. Loglinear models are also often used for cost analysis and may include efficiency scores as elements of the function, while Gibbs sampling is often employed to solve robust or ridge regression models. Each one of these methodologies is now discussed.

Data Envelopment Analysis and Deterministic Cost Models

Data Envelopment Analysis is a mathematical programming approach developed by Charnes, Cooper, and Rhodes (1978) from the work of Farrell (1957). DEA leads to a fractional programming problem, normally converted to a linear program for ease of solving. Efficiency is often evaluated as in the classical engineering sense of the ratio of outputs to inputs. The DEA measures used in this dissertation are generally unit invariant, because optimality is independent of the input and output measures used assuming that these are kept constant between Decision-Making Units (DMUs). As an optimization method, DEA does not rely on the traditional assumptions required of many other types of analysis such as regression. DEA requires no explicit identification of underlying relations between inputs and outputs and does not require weights to be assigned a priori. Decision-Making Units (DMUs), defined as the entities that convert inputs into outputs, are selected for efficiency revaluation. A full description of the technique is reserved for Chapter 3.

Several studies have applied DEA in the military medical departments. Charnes, Cooper, Lieck, Assad, Golany, and Wiggins (1985) conducted arguably the first Data Envelopment Analysis in a military health care facility. These authors evaluated the efficiency of 24 Army military hospitals using criteria that are still relevant for inclusion in this analysis. The authors selected traditional workload criteria for analysis of outputs including personnel trained, relative work product, and clinic visits. These outputs are considered traditional elements of production in health care and are relevant for inclusion along with other less traditional factors. For inputs, the study evaluated Full Time Equivalent (FTE) employees by specific category, inpatient expenditures, outpatient expenditures, weighted procedures, occupied bed days, and operating room hours[2]. Despite the fact that the research was conducted 20 years previously, most of the variables included retain relevance for measuring the traditional workload functions, although the paradigm in military health care has shifted towards prevention and health promotion instead of treatment Most impressively, a training output is specifically included in this study, although prevention, readiness, and other aspects are absent, as they were less relevant measures in the 1980s.

Another DEA study of interest was conducted by Mihara (1990). Mihara's efficiency analysis of the utilization of personnel at Navy Medical Treatment Facilities using 1987-1988 data provided implications for resource allocation. In this study, Mihara initially employed DEA to provide efficiency scores pertaining to the utilization of personnel at individual U.S. Navy hospitals. Efficient facilities were then further analyzed using least squares methods to baseline physician requirements (which were deemed workload and beneficiary dependent) and professional staff requirements

---

[2] Occupied bed days, weighted procedures, and operating room hours were seen as status indicators and thus modeled as inputs, which are reasonably associated with the Relative Weighted Product output.

(which were deemed physician dependent). "In other words, the optimal composition of personnel in terms of output can be determined from the structural equations of hospitals that are efficient." This study reveals that DEA methodologies might be used in conjunction with other tools to provide implications for resource allocation. Mihara's work, while relevant, was primarily driven by raw workload statistics. While workload is an important aspect for resource allocation, it is not the only input or output to be considered. Readiness, prevention, training, and prevention measures are important as well.

Ozcan and Bannick (1994) used DEA to study trends in Department of Defense hospital efficiency from 1998-1999 using 124 military hospitals and data from the American Hospital Association Annual Survey. In a 1995 study, these authors also compared DoD hospital efficiency with that of Veteran's Administration hospital efficiency (n=284) using 1989 data. These studies were conducted at the strategic level under a different operational paradigm, prior to the large-scale adoption of managed care.

Coppola (2003) conducted a DEA study of military hospital facilities using 1998-2002 data. In his study, he selected the following input variables: costs, number of beds in the MTF, FTEs, number of services offered. For output variables, he evaluated surgical visits, Ambulatory Patient Visits (APVs), emergency visits, Case Mix Adjusted Discharges (CMAD), and live births. This study is heavily focused on workload as the primary measure for efficiency, a point of view not fully congruent with the current operations of military hospitals. Further, his study does not standardize workload as visits are segregated only by clinic type instead of by the Resource-Based Relative Value Units (RVUs). [Note: RVUs standardize outpatient workload by accounting for resource consumption (complexity). The Centers for Medicare and Medicaid, formerly known as the Health Care Financing Administration, established RVUs as a method to equitably reimburse physicians. See the Centers for Medicare and Medicaid, http://www.cms.hhs.gov/, for more information.]

DEA may be used in conjunction with other tools of analysis. In their study of public secondary schools in Texas, Arnold, Bardhan, Cooper, and Kumbhakar (1996), illustrate how DEA may be coupled with traditional Ordinary Least Squares analysis of loglinear models to produce satisfactory efficiency estimations. In this study, the authors show that the OLS regression and SFA do not provide results consistent with economic theory or expectations, because they deal with "central tendency" estimates without allowing for differences in efficient and inefficient performers. DEA is then employed to determine efficient public secondary schools in Texas. Subsequently, a dummy variable reflecting efficient versus inefficient schools is incorporated into OLS regression models. The results illustrate that the combined methodology approach produces results consistent with economic theory and successfully combines estimation for efficient and inefficient behavior as identifiable components in one model.

Similarly, Brockett, Cooper, Kumbhakar, Kwinn, and McCathy (2003) employed the same combined DEA and OLS methodology in evaluating advertising programs for military recruitment. The authors evaluated a “service specific” program for advertising in comparison with a “joint program.” Using data from a previously conducted “designed experiment” advertising study, the authors showed that joint recruitment efforts are less efficient then service specific recruiting. (This approach will be employed later in various models in this research.)

Another interesting and relevant DEA study was conducted by Brockett, Chang, Rousseau, Semple, and Yang (2003). These authors evaluate Health Maintenance Organizations, which employ Independent Practice Associations (IPA) versus those that employ group/staff arrangements in a “game-theoretic” DEA model. In this model, the authors combine the two-person zero sum game approach with DEA, evaluating the results from both society’s and the consumers’ perspectives. Individual DMUs from one group are compared to the collective second group (or the efficient frontier from the second group). This technique is relevant when there are components of a system that may be in competition with each other. Specifically, the civilian network component of the military health care system versus the MTF components might be evaluated using this unique DEA approach.

Several relevant statistical cost models for hospitals have been developed by previous researchers. Cremieux and Ouellette (2001) proffered the following log-linear model:

$$(C_h) = \alpha + \beta_{ih}(V) + \gamma_{ih}(IP) + \delta_{ih}(K) + \zeta_{ih}(T) + \lambda_{ih}(O) + \xi_{ih}(T) + \varepsilon \qquad (1)$$

where
C_h = Cost function of the hospital
V = Volume / output measures (e.g., workload)
IP = Input price (normally labor rates)
K = Quasi-fixed / capital input vector (e.g., FTEs)
O = Hospital specific variables (e.g., profit vs. not for profit)
T = Technological change vector
and α, β, γ, ζ, λ, and ξ are the coefficients to be estimated with ε representing random statistical error.

West (2000) modified the above model to apply to the Army medical community in unpublished research. His log-linear model follows.

$$(Cost_{ht} \mid Inflation) = \alpha + \beta_{ht}(V) + \zeta_{ht}(X) + \lambda_{ht}(C) + \xi_{ht}(U) + \varepsilon \qquad (2)$$

where:

V = Volume / output measures (e.g., inpatient admissions and outpatient visits)
X = Complexity variables (casemix and GME)
C = Capacity factors (lagged admissions and outpatient visits)
U = Demand uncertainty (Coefficient of Variation for C by year)
and the parameters to be estimated are α, β, ζ, λ, and ξ.

The volume measures selected by West included admissions, outpatient visits, and length of stay. The complexity variables included the number of services, the case mix index, and a binary variable to detect Medical Center status. West derived measures for utilization of inpatient and outpatient services and used the coefficient of variation for determining demand uncertainty. He also used interaction terms and lagged variables. Most interestingly, West uses the cost model above to generate efficiency analysis by calculating the following: (Actual cost – Estimated cost) / (Estimated cost). His results are interesting in that the surviving variable groups are *V* and *X* only. Specifically, the only significant coefficients (at the .05 level or less) from the pooled OLS regression models include the interaction term between admissions and visits, case mix index and inpatient scope of work, lagged mean admits, and the categorization of facility as a MEDCEN. These variables suggest that a combination of outpatient and inpatient visits should be used for efficiency analysis.

Stochastic Cost and Efficiency Studies

Stochastic Frontier Analysis (SFA) applied to health care and other settings are detailed in this section. It is interesting to note that no SFA studies are on file with the Defense Technical Information Center (DTIC), which indicates that few if any applications of such models have been performed.

Stochastic Frontier Analysis is a relatively new body of work, which assumes that cost may be modeled as a function of variables with an associated error term that has two parts: noise and efficiency. SFA was first introduced in 1977 by Meeusen and Van den Broek and by Aigner, Lovell, and Schmidt (1977).

The SFA noise variable is often assigned a normal distribution, and the efficiency variable is often assigned a half-normal, exponential, or truncated normal distribution such that the total cost of any given hospital is a function of the known inputs and outputs plus the two-part error term. This leads to what is called the "composed error" model which assumes that the Ordinary Least Squares single error variable is actually composed of two random variables, one related to noise and the other related to inefficiency. Estimates for both efficiency and random noise derive from the solution to SFA models. A detailed discussion of the methodology is provided in Chapter 3.

One of the first applications of SFA to medical facilities (if not the first) was performed by Hofler and Folland (1991). Hofler and Folland suggest that SFA is important in assessing hospital costs and efficiencies, because other methods do not necessarily identify what minimum costs should be. The authors suggest that DEA is not entirely satisfactory because it ignores random fluctuations present in the data observations. In their research, Hofler and Folland assume structural cost differences based on ownership (for profit or not for profit categorization), teaching status, metropolitan or rural categorization, and Medicare volume ratio (high, low). The authors determined that inefficiency was responsible for about 10.5% of hospital costs overall. The number of cost equations (12 equations based on the assumed differences in cost structure) and the irregularity of group size (as small as 35 and as large as 442) served to illustrate some of the problems associated with SFA: several equations could not be estimated, as the Maximum Likelihood Estimates did not converge. A total of 422 hospitals were eliminated from their analysis. Further, across group comparisons were not possible.

Zuckerman, Hadley, and Iezzoni (1993) employed Stochastic Frontier Analysis for hospital cost functions as well. Specifying a cost model that relied on input prices, output volumes, and output characteristics (i.e., less tangible output results such as the

Joint Commission on the Accreditation of Health care Organizations score), the researchers determined that inefficiency accounts for (on average) 13.6 percent of total hospital costs, a result similar to Hofler and Folland.

William Greene (2003) recently used SFA in analyzing the World Health Organization (WHO) data regarding health care efficiency by country. He illustrated that WHO's efficiency analysis confused heterogeneity with inefficiency. His results differed substantially from the WHO findings, which demonstrated that the reliance on only a single methodology might mislead decision makers.

An interesting and relatively recent SFA study was conducted by Gerdtham, Loethgren, Tambour, and Rehnberg (1999). These authors developed a multiple output, multi-input SFA model and applied it to the Swedish health care system. The authors used a stochastic "ray translog production function," which provides the maximum norm of the attainable outputs. Using the ray norm of the number of operations, number of visits, and number of discharges, the authors evaluate the effects of reimbursement schemes and conclude that 9.7% might be saved by adopting a workload based method of financing county hospitals versus a budget based model. While the model was focused strictly on workload, the inclusion of multiple outputs made the research unique and noteworthy.

Bayesian studies of cost and efficiency in the military healthcare setting are not available in the Defense Technical Information Center (DTIC), which suggests that few or none have been conducted. A recent manuscript by Kumbhakar (2003) used

Bayesian techniques to estimate input-output production efficiency. Instead of using the traditional composed error model associated with SFA, Kumbhakar evaluated a nonlinear mixture model requiring empirical methods for sampling. Similarly, this dissertation uses empirical sampling of mixture models for ridge and robust regression.

Synopsis

From this review, it becomes clear that deterministic and stochastic models have been applied extensively to research in the health care. Specifically, several DEA and stochastic models using a variety of models have been applied to this field. From these studies, it is also clear that most maintain a strictly production-oriented focus, and few efforts have applied both methodologies to the same study in order to compare results. This research proceeds in this vein. With a basic understanding of previous research in hand, one may now turn to the technical discussion of methods.

Inter alia, this dissertation deals with the bias that can accompany the use of only one method of analysis, which is called "methodological bias" by Charnes, Cooper, and Sueyoshi (1988). Cross checking with other methods provides added insight into the accompanying results. The importance of cross checking with different methodologies is exhibited in a study of North Carolina hospitals by Bauer, Conrad, and Straus (1986), which contrasted the use of statistical regression (with translog cost models) versus DEA. While the translog models universally found constant returns to scale to prevail, the DEA model found increasing and decreasing returns to scale in nearly all of the individual hospitals. Thus, the use of multiple methodologies is important for decisions regarding allocation of scarce resources. The possible methods to be used are discussed in the next chapter.

Chapter 3: Methods

Data Envelopment Analysis

Data Envelopment Analysis is a flexible, mathematical programming approach for the assessment of efficiency, where efficiency is (in general) defined as a linear combination of the weighted outputs divided by a linear combination of the weighted inputs. DEA models are flexible, albeit primarily deterministic unless some sort of stochastic modifier is used. A detailed development of several DEA models follows.

There are multiple formulations of DEA models. Two of the most frequently used models are the Charnes, Cooper, and Rhodes (1978) model (CCR) and the Banker, Charnes, and Cooper (1984) model (BCC). The CCR model assumes constant returns to scale, which is arguably less appropriate for evaluation of health care entities, because there should be no assumption of constant returns to scale (Zuckerman et al, 1994). Constant returns to scale implies that if $f(x) = y$ represents the production function, f, which relates output vector y to input vector x, then $f(tx) = t(f(x)) = ty$, where t is any positive scalar (Cooper, Seiford, and Tone, 1994).

Charnes, Cooper, and Rhodes Models (CCR)

Assume that an organization wishes to assess the relative efficiencies of some set of comparable subunits. (The subunits are called Decision Making Units or DMUs.) Further assume that management believes that the production frontier represents a function that exhibits constant returns to scale. For each DMU, there is a vector of associated inputs and outputs of managerial interest. (Determining an appropriate set of outputs and inputs is critical and requires much care.) In this case, the manager is interested in either maximizing the outputs while not exceeding current levels of inputs (output oriented) or minimizing the inputs without reducing any of the outputs (input oriented). Assume that the input orientation is preferred. The manager assumes that the traditional definition of engineering efficiency (ratio of weighted outputs to weighted inputs) will result in an acceptable solution for technical efficiency. With these assumptions in place, one may formulate the following fractional programming problem that may be solved to determine technical efficiency, defined (for now) as the ratio of weighted outputs to weighted inputs, for each separate DMU:

$$\text{Maximize Efficiency for DMUo} = \theta = \frac{\sum_{i=1}^{m} u_i y_{io}}{\sum_{j=1}^{n} v_j x_{jo}} \qquad \text{(A)}$$

$$\text{Subject to}: \frac{\sum_{i=1}^{m} u_i y_{iz}}{\sum_{j=1}^{n} v_j x_{jz}} \leq 1, z = 1,2,....n \qquad \text{(B)}$$

$$u, v \geq 0$$

In this formulation, there are *m* outputs, *n* inputs, and *z* Decision Making Units (DMUs). Efficiency is designated as θ. The output data y_{io} is the value of output *i* for DMU o, while x_{jo} is the value of input *j* for DMU *o*. (Note: this mathematical program is run *z* times, once to determine the efficiency of each DMU. That is, while statistical regression optimizes across all observations, DEA optimizes on each observation.) The index *o* identifies the selected DMU for which an efficiency score will be generated. The components of the vectors [*u,v*] are the weights to be determined for the *m* outputs and *n* inputs respectively. This model defines efficiency for the selected DMU as the weighted linear combination of its outputs divided by the weighted linear combination of its inputs, subject to the constraint that, for each DMU (including the one whose index *z* is *o)*, the efficiency cannot exceed one. All weights are restricted to be nonnegative.

This formulation is nonlinear; however, straightforward transformations convert this formulation to a linear program. The ratio in the objective can be solved by maximizing the numerator subject to the condition that the denominator equals one, or by minimizing the denominator, subject to the constraint that the numerator equals one. To do the former, one defines the variable

$$t * \sum_{j=1}^{n} v_i x_{jo} = 1 \tag{C}$$

which implies $t > 0$. Multiplying (A) by this new variable, (C), from the previous equation results in a revised objective function $Maximize\ t\sum_{i=1}^{m} u_i y_{io} \rightarrow \sum_{i=1}^{m} \mu_i y_{io}$, where μ_i is nothing more than the scalar *t* multiplied by each value of u_i. The final linear programming problem follows (Cooper et al, 2000).

$$\text{Maximize } \mathrm{E} = \sum_{i=1}^{m} \mu_i y_{io} \tag{3}$$

Subject to:

$$\sum_{i=1}^{m} \mu_i y_{io} - \sum_{j=1}^{n} v_j x_{jo} \le 1, z = 1,2,\ldots.n$$

$$\sum_{j=}^{n} v_i x_{jo} = 1$$

$$u,\ v,\ \mu \ge 0$$

From inspection, it is clear that the production possibility set is nothing more than a cone, noting that $C \in \Re^n$ is a cone if $\lambda \boldsymbol{x} \in C\ \forall\ \lambda > 0,\ \boldsymbol{x} > 0$ where $\boldsymbol{x}$ are the elements of the cone (Bertsimas and Tsitsiklis, 1997). Table 2 provides a comparison of (3) and its associated dual along with the output oriented CCR model.

Table 2. CCR Models.

The primal and the dual of the CCR models are provided side-by-side (Cooper et al, 2004).

Input Oriented Multiplier Form

Maximize $\sum_{i=1}^{m} \mu_i y_{io}$

Subject to:

$$\sum_{i=1}^{m} \mu_i y_{io} - \sum_{j=1}^{n} v_j x_{jo} \le 1, z = 1,2,....n$$

$$\sum_{j=}^{n} v_i x_{jo} = 1$$

$u, v, \mu \ge 0$

Input Oriented Envelopment Form

Minimize $\theta - \eta(es^{-}_{D} + es^{+})$

Subject to:

$$Y\lambda - s^{+} = y_o$$

$$X\lambda + s^{-} = \theta x_o \qquad (4a)$$

$\lambda \ge 0, s^{+} \ge 0, s^{-} \ge 0$

Output Oriented Multiplier Form

Minimize $\sum_{i=1}^{n} v_i x_{io}$

Subject to:

$$\sum_{i=1}^{n} v_i x_{io} - \sum_{j=1}^{m} \mu_j y_{jo} \le 1$$

$$\sum_{j=1}^{m} \mu_j y_{jo} = 1$$

$v, \mu \ge 0$

Output Oriented Envelopment Form

Maximize $\theta + \eta(es^{-}_{D} + es^{+})$

Subject to:

$$Y\lambda - s^{+} = \theta y_o \qquad (4b)$$

$$X\lambda + s^{-} = x_o$$

$\lambda \ge 0, \theta \ge 0$

where y_o is the column vector of outputs for DMU_{o}, u^T is the transposed vector of output multipliers (so that $u^T y_o$ is a scalar), x_o is the vector of inputs for DMU_{o} and Y and X are matrices of outputs and inputs respectively.

To provide clarity, an illustration with one input and one output is provided. Suppose that one wishes to analyze the technical efficiency of Army hospitals using the input oriented CCR model (4a), using normalized outpatient visits as the single output and budgeted dollars as the single input. In this case (and using actual data), the feasible set is a cone in $\Re^2$ as illustrated in Figure 4.

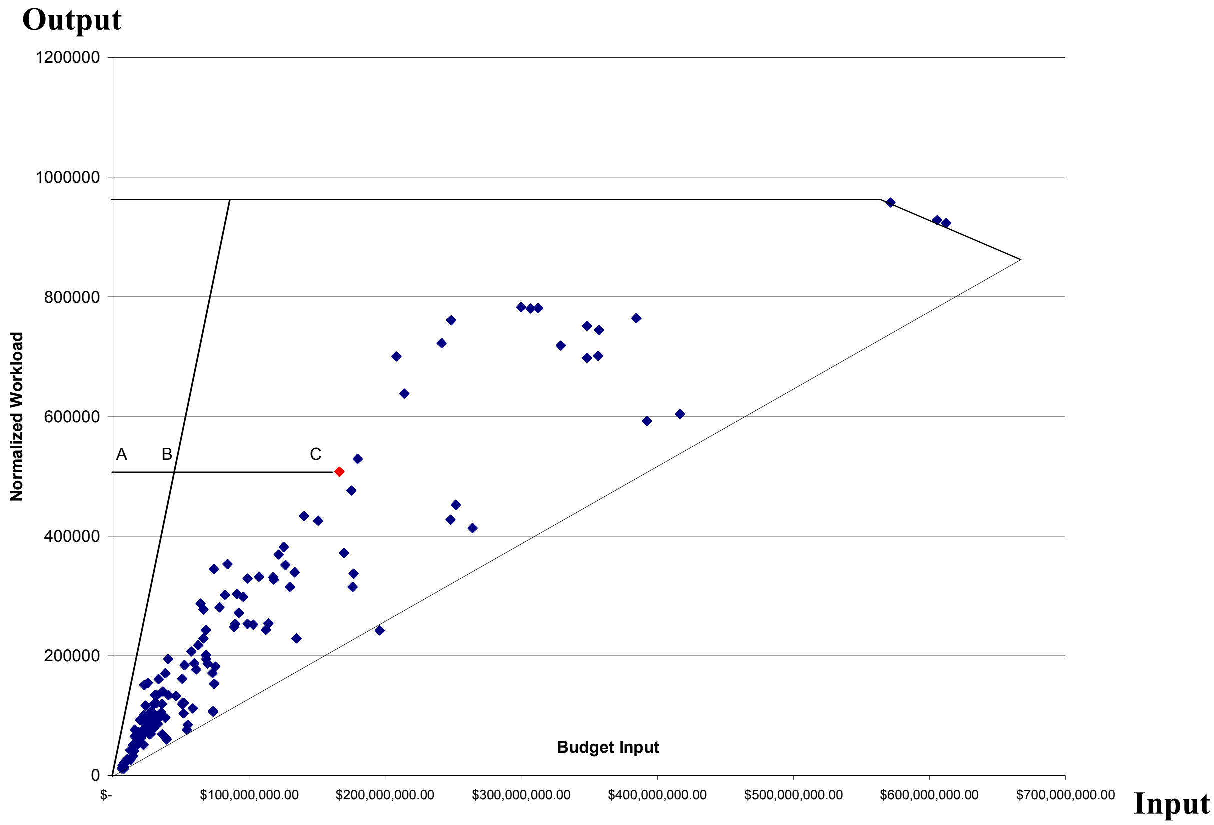


Figure 4. CCR Cone, One Input and One Output.

The production possibility set for this two-dimensional CCR model is a cone. The efficient frontier is the ray from the origin through the DMU with the maximum output. The ratio (A:B) / (A:C) is the input oriented CCR efficiency score for the selected DMU. One can see that this model is susceptible to outliers, as it must proceed through any outlying points. Further, the CRS assumption implies that the relationship between budget and workload should be linear.

Consider the addition of another input, Full Time Equivalent (FTE) personnel. Then the production possibility set is a three-dimensional cone as depicted below.

.

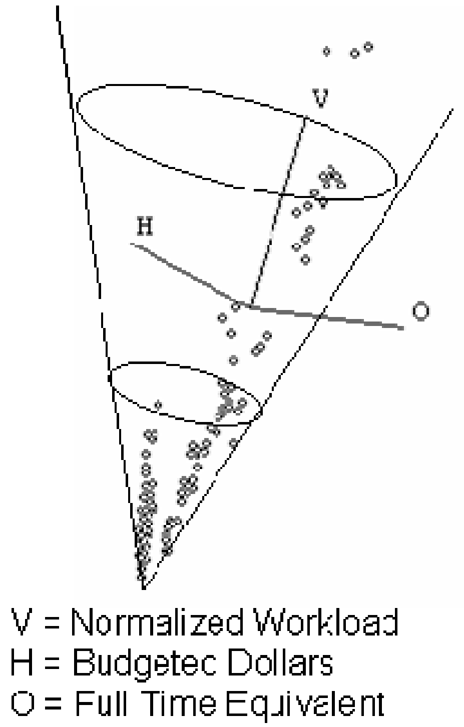


Figure 5. CCR Cone, One Input and Two Outputs.

The addition of FTEs is shown above (with some rotation for clarity). Again, the region is a cone. At each cross-sectional point, efficient units are the extreme points on the extreme rays. (The depiction of circular cross-sections is convenient albeit inaccurate.) Again, one can see that an entity that appears to be extremely efficient in any single input axis may vastly affect the efficiency scores of the other entities.

While these formulations of the CCR model provide an efficiency score (specifically, the optimal value for θ defined as θ^* in the solution to the dual problem in (1)), there is no allocative efficiency obtainable from this one output-one input formulation. If the optimal θ defined as θ^* is equal to one ($\theta^*=1$), the evaluated DMU is measured as efficient; however, if there are multiple outputs and multiple inputs, it might be possible that a reduction of input or an increase in output is possible even with $\theta^* = 1$. Consider adding an output for weighted inpatient visits and removing the input variable FTE. In the output oriented case, the plot of the normalized outpatient workload per budgeted dollar versus normalized inpatient workload per dollar reveals that the production possibility set has technically efficient hospitals; however, the allocation might be adjusted. A DMU might be able to produce more output while maintaining inputs constant and is therefore not fully efficient (see Figure 6).

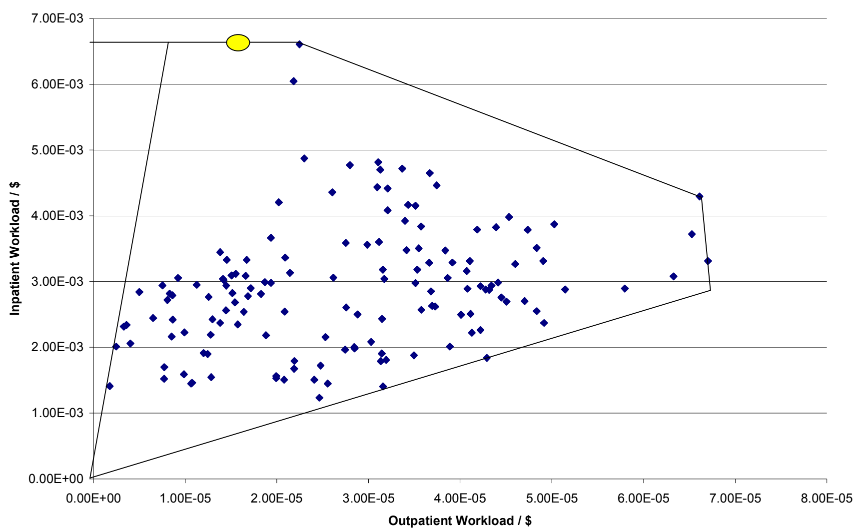


Figure 6. Allocatively Inefficient DMU.

This two output model illustrates the difference between technical and mix efficiency. Specifically, the larger point, while Farrell efficient, produces less outpatient workload per budgeted dollar than the point immediately to its right and is therefore allocatively inefficient.[3]

The solution to this form of the allocative efficiency problem requires a two phase approach. Phase 1 solves the DEA linear program and obtains an optimal theta

[3] This condition is also called "weakly efficient," because it lies on the line at the top of Figure 6, which is not part of the efficient frontier (given by the slanted lines from the origin and on the left side of the figure.

for the DMU being evaluated. In Phase II, the optimal objective function value, θ^*, obtained from Phase I is used to determine the maximum slack solution, where the input and output slacks are defined as follows:

$$\begin{array}{lll} -X\lambda + \theta x_o \geq 0 \Rightarrow -X\lambda + \theta x_o - s^- = 0 \rightarrow s^- = \theta x - X\lambda & & (5) \\ Y\lambda \geq y_o \Rightarrow Y\lambda - s^+ = y_o \rightarrow s^+ = Y\lambda - y_o & & \end{array}$$

where s^- and s^+ are vectors of slack variables for the inputs and outputs respectively. If the maximum slack solution is zero, then the unit is deemed allocatively efficient. From (5), one formulates the Phase II linear programming problem as follows.

Maximize $-es^- + es^+$	*e is defined as a vector of 1's*
Subject to:	
$\theta^* x_o - X\lambda = s^-$	*Definition of input slack*
$y_o = Y\lambda - s^+$	*Definition of output slack*
$\theta^* \geq 0,\ \lambda \geq 0,\ s^- \geq 0,\ s^+ \geq 0$	*Non-negativity constraints*

where θ^* is the optimal value of θ determined in (4), e is a row vector will all elements unity and s^- and s^+ are the column slack vectors for input and output slack respectively.

A DMU that has $\theta^* = 1$ and has a zero-slack solution (for all slacks) is technically efficient or Pareto-Koopmans efficient. As defined in Cooper, Seiford, and Tone (2000), Pareto-Koopmans efficiency is attained only if it is impossible to improve any input or output without worsening some other input or output. This may be given precise mathematical form by the following definition. DMU_i (the DMU being evaluated relative to the performance of all DMUs) is technically efficient if and only if the following two conditions are both satisfied: (i) $\theta^* = 1$, (ii) All slacks are zero (allocative efficiency is achieved). In all other cases it is possible to improve one or more of the inputs or outputs without worsening any other input or output. A DMU that achieves (i) and (ii) is then called "Pareto-Koopmans" efficient (Cooper, Seiford, and Tone, 2000).

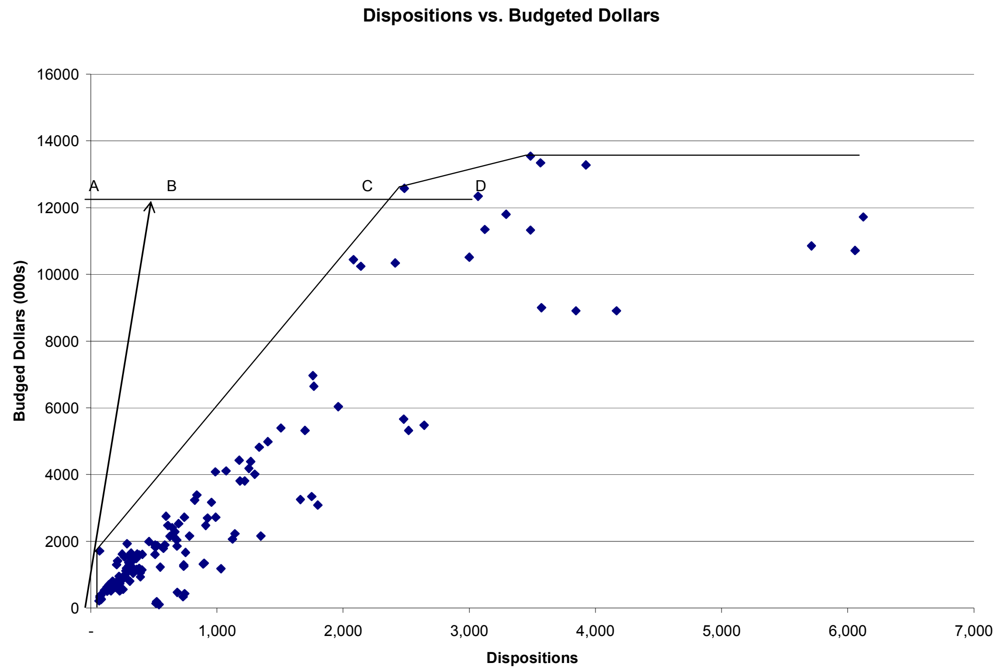
Banker, Charnes, and Cooper Models (BCC)

With the basic formulation of the CCR DEA model in place, it is now necessary to modify the mathematical program in order to meet the requirements of the health care setting. The assumptions of the simple CCR model are not fully appropriate for the medical community. Specifically, one might suspect that the relationship between inputs and outputs involves variable returns to scale, i.e., that there exists a varying relationship between increasing output and input. The BCC model must then be considered.[4]

Consider now, for example, hospital dispositions as an output versus budgeted dollars as an input. Figure 7 (derived from actual data) is suggestive of variable returns to scale. As the amount of budgeted dollars increases, one sees a "leveling off" so the output increases at a decreasing rate followed by a decrease in the output. The decreasing returns may be a result of improper budgeting, random noise, or the effects of other variables, or a combination of all the above and more.

[4] Investigation of the returns to scale in CCR models may be readily completed by evaluating the sum of $e\lambda^*$ at an optimum where $e\lambda^*=1 \Rightarrow$ constant returns to scale, $e\lambda^* < 1 \Rightarrow$ increasing returns to scale, $e\lambda^* > 1 \Rightarrow$ decreasing returns to scale (Banker, Cooper, Seiford, and Zhu, 2004).

Output



Input

Figure 7. BCC Formulation.

The figure above suggests that there exist variable returns to scale. The difference between the calculated efficiency under the constant returns to scale assumption (AB / AD) and the variable returns to scale assumption (AC / AD) is obvious and significant.

The BCC model improves the CCR model by introducing non-constant returns to scale. The mathematical program of the BCC model includes a single additional constraint; specifically, the sum of the dual multipliers must equal one (a convexity constraint). The formulations of the input oriented and output oriented BCC models follow (Cooper et al, 2004).

Input Oriented Multiplier Form

Maximize $\sum_{i=1}^{m} \mu_i y_{io} - \mu_o$

Subject to:

$$\sum_{i=1}^{m} \mu_i y_{iz} - \sum_{j=1}^{n} v_j x_{jz} - \mu_o \le 1, z = 1,2,....n$$

$$\sum_{j=}^{n} v_i x_{jo} = 1$$

$v,\ \mu \ge 0$

Input Oriented Envelopment Form

Minimize $\theta - \eta(es^{-}_{D} + es^{+})$

Subject to:

$$Y\lambda - s^{+} = y_o$$

$$X\lambda + s^{-} = \theta x_o \qquad (6a)$$

$$e\lambda = 1$$

$\lambda \ge 0, s^{+} \ge 0, s^{-} \ge 0$

Output Oriented Multiplier Form

$\sum_{i=1}^{n} v_i x_{io} + v_0$

Subject to:

$$\sum_{i=1}^{n} v_i x_{io} - \sum_{j=1}^{m} \mu_j y_{jo} + v_0 \le 1$$

$$\sum_{j=1}^{m} \mu_j y_{jo} = 1$$

$v,\ \mu \ge 0$

Output Oriented Envelopment Form

Maximize $\theta + \eta(es^{-}_{D} + es^{+})$

Subject to:

$$Y\lambda - s^{+} = \theta y_o$$

$$X\lambda + s^{-} = x_o \qquad (6b)$$

$$E\lambda = 1$$

$\lambda \ge 0, \theta \ge 0$

In the multiplier form, all variables are as in the CCR model except that the variable μ_o is included in the objective to be maximized. This variable is associated with the convexity constraint $e\lambda=1$ in the envelopment form. In the envelopment form, the previous two-step formulation is now combined, so that it is reflected in a single model. This is accomplished by introducing the scalar, η, as a value so small that it is smaller than any positive real number. This is accommodated logically by regarding θ has having "preemptive priority" so that no increase in the slack variable values can compensate for any decrease (no matter how small) in θ. See the previous discussion where the second step optimization with θ changed from a variable to a constant, θ^*, as determined from the stage one optimization.

Even as written, the BCC model is insufficient for analysis of the military health care system. Consider the fact that there exist "non-discretionary" inputs or outputs (depending on orientation). For example, a hospital probably does not influence the population demographics that it supports. These inputs are nondiscretionary. Further, there may be requirements to produce outputs, which are non-discretionary. An example of a non-discretionary output is the requirement for the maintenance of a deployable capacity in military hospitals. This readiness function is measurable in terms of budgeted dollars expended; however, the maximum amount that can be expended is non-discretionary.

Adjustments for non-discretionary inputs and outputs are relatively straightforward. Simple segregation of the input or output data into two sets (discretionary (D) and non-discretionary (ND)) are accomplished in the revision of (6a), which is (again) an input oriented formulation.

$$\begin{aligned}
&\text{Minimize } \theta - \eta(es^-_D + es^+) \\
&Y\lambda - s^+ = y_o \\
&X\lambda + s^-_D = \theta x_o \text{ for } X, x_o, s^- \in D \\
&X\lambda + s^-_{ND} = x_o \text{ for } X, x_o, s^- \in ND \\
&e\lambda = 1 \\
&x \geq 0, y \geq 0, \lambda \geq 0, \beta \geq 0
\end{aligned} \qquad (6c)$$

This formulation partitions the inputs and input slacks (s^-) into two mutually exclusive and categorically exhaustive sets, discretionary (*D*) and non-discretionary (*ND*). One can readily see that the non-discretionary input slacks are not included in the objective function and do not enter the measure of efficiency evaluation that is being obtained. Further, they are not multiplied by θ in the constraint set, so the nondiscretionary input may not be reduced. A full discussion with accompanying examples is provided in Cooper, Seiford, and Tone (2000, pg. 64).

The solution to both the CCR and BCC DEA linear programs produce useful results. Primal constraints, which are active (zero slack) produce the reference set for each DMU. Further, the Lagrange multipliers for the primal constraints are useful in sensitivity analysis. These multipliers provide the information necessary to make resource allocation recommendations if used appropriately, as will be discussed for each analysis.

<u>Non-Oriented Models (ADD, SBM, and RAM)</u>

Returning to the definition of technical efficiency, one can see that the condition (i) may be satisfied without satisfying condition (ii) of all slacks zero. Thus if only (i) is satisfied, the result is referred to as "weakly" efficiency. If both (i) and (ii) satisfied, the DMU is "fully" or "strongly" efficient.

Several efficiency formulations focus on the analysis of slacks. These non-oriented formulations may prove useful in developing resource based decisions. The information derives from Cooper, Seiford, and Tone (2000).

The basic additive model (or ADD) is formulated as follows.

$$
\begin{aligned}
&\textit{Maximize } es^{-} + es^{+} \\
&Y\lambda - s^{+} = y_o \\
&X\lambda + s^{-} = x_o \\
&e\lambda = 1 \\
&\lambda \geq 0,\ s^{-} \geq 0,\ s^{+} \geq 0
\end{aligned}
\tag{7}
$$

In this formulation, the objective function is to maximize the input and output slacks subject to constraints that the sum of all the weighted outputs (inputs) plus the output (input) slacks must equal the output (input) of the DMU under investigation. A convexity constraint is also included for VRS (Cooper, et al, 2000). The addition of non-discretionary variables is simple. The appropriate variable slack is simply removed from the objective function. While this method does not provide a single measure of efficiency, it does focus on “full” efficiency via the following definition.

Definition 2: Efficiency is attained in (7) if and only if all slacks are zero at the optimum.

The Slacks Based Measure (or SBM) is a units invariant approach that assumes all the inputs and outputs are positive and may readily expressed as a fractional programming problem as follows.

$$Minimize\ \rho = \frac{1 - \sum_{j=1}^{m} \frac{s_j^-}{x_{jo}}}{1 + \sum_{k=1}^{s} \frac{s_k^+}{y_{ko}}} \text{ for each DMU o}$$

$$\begin{aligned} &Y\lambda - s^+ = y_o \\ &X\lambda + s^- = x_o \\ &e\lambda = 1 \\ &\lambda \geq 0,\ s^- \geq 0,\ s^+ \geq 0 \end{aligned} \qquad (8)$$

In this fractional formulation, the objective function ρ ranges between zero and one and is a function of the "ratio of slack ratios." The objective function,ρ, then provides a unidimensional measure based on slacks for specifying efficiency (Cooper, et al, 2000).

Another non-oriented DEA model is known as the Range Adjusted Measure or RAM model, which is also units invariant but does not require the data to be positive definite or even semi-positive definite. The basic model is similar to (7) and is expressed as follows.

$$Maximize \sum_{j=1}^{m} \frac{s_j^+}{R_j^+} + \sum_{k=1}^{n} \frac{s_k^-}{R_k^+}$$

$$\begin{aligned} &Y\lambda - s^+ = y_i \\ &X\lambda + s^- = x_i \\ &e\lambda = 1 \\ &\lambda \geq 0,\ s^- \geq 0,\ s^+ \geq 0 \end{aligned} \qquad (9)$$

where R_j^+ is the range from the lowest to highest values observed for output j (j= 1….m), while R_k^- is the range from the lowest to the highest values observed for input k (k = 1…n).

This model is identical to the additive model, (7), in the constraints; however, the objective function has been modified to normalize the slacks based upon the range R of the inputs and outputs. An advantage to this derivation is that the result is now translation invariant in the direction of both inputs and outputs. See Cooper et al (2000, pg. 228) for the proof.

Comparison of Attributes for CCR, BCC, ADD, SBM, and RAM Models

Every formulation discussed has limitations that should be evaluated (see Table 3). In this dissertation, there is a requirement for location transformations (translation) in the linear models associated with the output direction. (These transformations are to ensure that data are strictly positive prior to logarithmic transform). Therefore, it would be useful to evaluate models that are translation invariant in the outputs. Further, reliable estimations of efficiency require unit invariance, a property that ensures the objective function is invariant to changes in the measurement units of the inputs and outputs. Next, there is a requirement for at least some of the DEA formulations to produce a single measure of efficiency. Further, formulations should generally be VRS based upon knowledge of the healthcare field. The following table suggests that the best of the methods discussed are the BCC-I and the RAM models.

Table 3. Model Characteristics.

The following table provides a synopsis of the model properties discussed. Only the BCC-I and RAM models are compliant. This table is adapted from Table 4.4 of Cooper et al (2000) but includes an additional column for the RAM characteristics.

Model Characteristics (adapted from Cooper et al 2000, p. 102)

		CCR-I	**CCR-O**	**BCC-I**	**BCC-O**	**ADD**	**SBM**	**RAM**
Data	Inputs	Semi-pos	Semi-pos	Semi-pos	Free	Free	Free	Free
	Outputs	Free	Free	Free	Semi-p	Free	Free	Free
Translation Invariance	Inputs	No	No	No	Yes	Yes	No	Yes
	Outputs	No	No	Yes	No	Yes	No	Yes
Units Invariance		Yes	Yes	Yes	Yes	No	Yes	Yes
Efficiency Score		[0,1]	[0,1]	(0,1]	(0,1]	No	[0,1]	Yes
Returns to Scale		CRS	CRS	VRS	VRS	BOTH	BOTH	BOTH

Window Analysis

Window analysis (Cooper et al, 2004) is a simple method for analyzing DMU efficiency over time using the DEA methodology. In this formulation, each DMU at time *t* is considered a separate entity (e.g., DMU_{a1}, DMU_{a2}, DMU_{a3}). Time periods are grouped (e.g., groups of three month intervals) and efficiency scores are calculated for DMU entries in that timeframe. One group (by date) of DMUs is then dropped and a new group of DMUs is added in a manner reminiscent of the way moving averages are calculated, and the process and evaluations are repeated. Because of the few number of time periods associated with the dataset in this dissertation, window analysis is not used. Capturing changes over time, however, is critical to model validity and is addressed by adjustments for inflation, grouped analysis, and ungrouped analysis. Specifically, this dissertation considers each hospital *i's* performance at year *t* a separate

DMU. Efficiency comparisons are then made for $i*t$ DMUs and are then segregated for further analysis by time period t. This approach allows for larger differentiation and more degrees of freedom in evaluating efficient and inefficient performers by increasing the number of DMUs per inputs and outputs.

Capabilities and Limitations of DEA

DEA, like any modeling tool, has strengths and weaknesses. A synopsis of these strengths and weaknesses is provided in Table 4. By itself, DEA is descriptive in nature; however, slack and sensitivity analysis provide certain insight into potential allocation shifts. When coupled with linear models, the results of DEA have been shown to improve overall model performance (Arnold et al, 1996). The next section develops this procedure more fully.

One of the weaknesses (partially addressed by stochastic models) is the sensitivity of DEA with regards to outliers. Superefficient entities may skew the results. As with any analysis, careful graphing and evaluation of the data is critical and certain stochastic methods might be employed to reduce the impact of noise on efficiency analysis, which will eventually lead to the discussion of stochastic models. Various tools have been developed to identify "outliers" and other affects of DEA. This includes the use of "efficiency matrices" like that depicted on page 37 of Cooper, Seiford, and Tone (2000) that shows how each DMU is used either to evaluate other DMUs or to be evaluated by them.

Table 4. Strengths and Weaknesses of DEA.

A listing of the most relevant strengths and weaknesses of DEA is provided below.

Strengths	***Weaknesses***
Unit invariance: units for inputs and outputs irrelevant	Sensitive to outliers
No requirements for data normality, independence, or identical distribution	Orders achieved efficiency, but this may not equate to attainable performance
Provides measurement of slacks and method for sensitivity analysis	Efficiency scores are sensitive to sample size
Does not require specification of functional form (like statistical models including regression require)	
Does not require pre-specified weights	

Linear Models

When the results of DEA are used as variables in linear models, it is possible to produce estimates for use in resource allocation as illustrated by Mihara (1990). This next section provides development of the traditional linear model and the inclusion of DEA results in such models. A discussion of Corrected Ordinary Least Squares is also provided.

A general purpose of linear statistical models is to test the relationship between one or more dependent variables with independent variables and covariates with the assumption that the parameter values may be estimated by linear models such as the normal equations used to estimate the parameters in Ordinary Least Squares (OLS). The term "linear" refers only to the way the parameters are estimated, not to the form of the function that relates inputs to outputs. The terms might be polynomial and include variable interactions, as in quadratic polynomials. Furthermore, the linear relationship of the parameters may be the result of transformations (e.g., logarithmic).

Simple Linear Models and Box-Cox Transformations

Simple linear models describe the relationship between a single dependent variable and independent variable and are expressed by the model

$$y|x = \beta_0 + \beta_1 x + \varepsilon \Rightarrow \qquad (10)$$
$$E(y|x) = \beta_0 + \beta_1 x,$$

where y is the response variable, x is the regressor term, ε is the error term (independent of x and generally assumed to be centered at zero with constant variance), $\beta = [\beta_0, \beta_1]$ is the parameter vector (intercept and slope) to be estimated, and $E(y|x)$ is the expectation of y given the values of x (Cook & Weisberg, 1999). The minimal set of assumptions for this model is that the above model reflects the true relationship of x and y, that the error terms are independently distributed with zero mean and constant variance σ^2, that the predicted mean function is truly linear (model accuracy), and that the observations were drawn independently. Gaussian assumptions are required for certain inferential tests. Only y is usually considered a random variable, and x is normally thought to be fixed, although there are methods for assuming that x is also random (Metzger, 2002). One immediately notes that the assumptions pertaining to the error imply that (10) is linear as

$$E(y|x) = E(\beta_0 + \beta_1 x + \varepsilon_i) = \beta_0 + \beta_1 x + E(\varepsilon_i) = \beta_0 + \beta_1 x, \text{ so } \hat{E}(y \mid x) = \hat{\beta}_0 + \hat{\beta}_1 x, \qquad (11)$$

where $E(y|x)$ refers to the expected value of y given the value of x and $E(\varepsilon) = 0$ on the assumption that the random error term ε is statistically independent centered at zero (and with constant variance). Further, the assumption of the constant variance of ε_i provides that

$$Var(\beta_0 + \beta_1 x + \varepsilon) = V(\varepsilon) = \sigma^2 I, \text{ so } \hat{V} = \hat{\sigma}^2 I \tag{12}$$

where $\hat{V}$ is the covariance matrix with non-zero entries only on the diagonal.

If the covariance matrix is diagonal but the elements are not identical, heteroskedasticity exists, and it is necessary to transform the response variable and regressor terms. If the observations are not drawn independently (e.g., time series), then $\hat{V}$ is a non-diagonal matrix, and the least squares solution includes this matrix, as will be discussed later. It is important that the model assumptions hold, so one needs to investigate the linearity of the assumed mean function, as illustrated below.

Assume that one would like to evaluate the relationship between dollars budgeted and normalized outpatient workload in a hospital environment through the linear model, then (10) might be applied as follows.

$$y|x = workload \mid dollars = \beta_0 + \beta_1 dollars + \varepsilon. \tag{13}$$

A quick scatterplot of sample data (Figure 8, which includes actual data for military hospitals in 2003) using a robust estimation of the mean function (a locally weighted scatterplot smoother or *lowess* estimate, see Cleveland, 1979) suggests that the mean function is non-linear. The associated standard deviation bands further suggest that the variance is not homoskedastic. Application of a linear model to such a relationship without transformation may be inappropriate, so further analysis is required to justify its use.

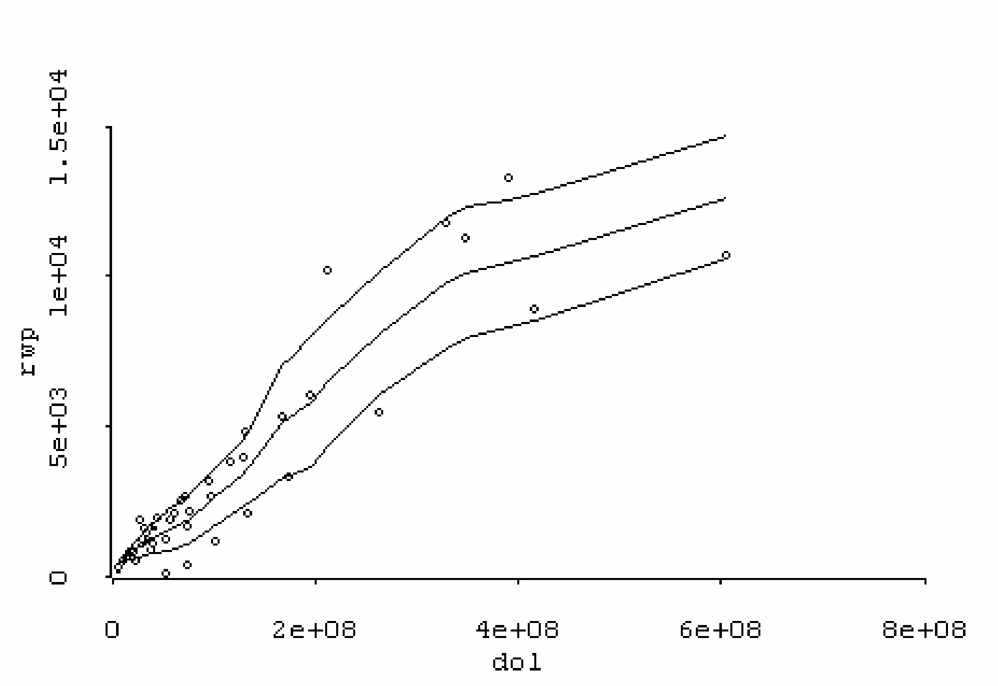


Figure 8. Heteroskedasticity, One Regressor Term.

The *lowess* estimate of the function in this simple, two-variable scatterplot reveals significant nonlinearity. (The label "rwp" on the vertical axis is the *workload* metric (Relative Weighted Product), while "dol" is *dollars* (or total facility costs) as reflected on the horizontal axis.) The standard deviation bands associated with the graph suggest increasing variance with increasing values of dollars, which implies heteroskedasticity of the variance. The variance is not constant as was assumed in (12). Simple regression on the untransformed data therefore appears to be inappropriate.

From above, it is clear that the transformation of the data is important to ensure that the assumptions of the linear model are met as much as possible. In this case, one might search for the appropriate transformation of both variables from a family of transformations using the Box-Cox method. (Other options would include the investigation of weighted linear models or nonlinear alternatives). The data suggest that a power transformation (possibly logarithmic) might be appropriate as there is a nonlinear relationship. In this case, one seeks to find the optimal λ for $y^{(\lambda)}$ and $x^{(\lambda)}$ where, for any random variable z,

$$z^{(\lambda)} = \begin{cases} \dfrac{z^{\lambda}-1}{\lambda} \text{ if } \lambda \neq 0 \\ \log(z) \text{ if } \lambda = 0 \end{cases} \tag{14}$$

.

This set is designed such that $\lim_{\lambda \to 0} \dfrac{z^{\lambda}-1}{\lambda} = \log(\lambda)$. Values of λ are found using a nonlinear optimization algorithm that minimizes the residual sum of the squared error in the modified power transformation model $z^{(\lambda)} gm(z)^{1-\lambda} \mid x = \beta_0 + \beta_1 x + \varepsilon$, where *gm(z)* is the geometric mean of the observed values of *Z*, which standardizes the units of Residual Sum of Squares (RSS) for all values of λ (Cook & Weisberg, 1999). Specifically, the nonlinear objective function may be stated as follows.

$$\underset{\lambda}{Minimize} \left(z^{(\lambda)} gm\left(z^{1-\lambda}\right) - \beta_0 - \beta_1 x\right)^2$$

$$\text{subject to: } z^{(\lambda)} = \begin{cases} \dfrac{z^{\lambda}-1}{\lambda} \text{ if } \lambda \neq 0 \\ \log(z) \text{ if } \lambda = 0 \end{cases} . \tag{15}$$

Each variable z_i may be investigated separately, or a joint minimization of the λ*'s* might be pursued.

Joint minimization using nonlinear methods results in the improved linear mean function for normalized outpatient workload versus budgeted dollars as seen in Figure

9. For the *workload* variable, the optimal parameter estimate is $\lambda = .28$, while for *dollars*, the best parameter estimate $\lambda=.17$ (close to the natural logarithm, which one might expect). Inferential test statistics and distributions are discussed later; however, a more reasonable model may be

$$log(workload \mid dollars) = \beta_0 + \beta_1 * log(dollars) + \varepsilon \Rightarrow$$

$$E(log(workload \mid dollars)) = \beta_0 + \beta_1 * log(dollars).$$

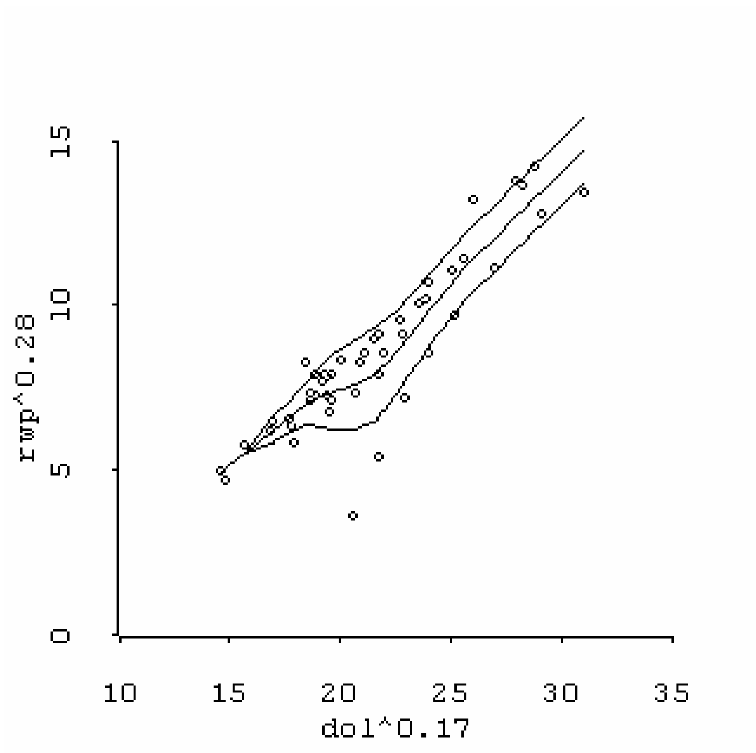


Figure 9. Box-Cox Transformation, One Regressor.

The estimated mean function is much improved after transformation. One might note that the variance bands, while imperfect, are also much improved. The assumptions for the relationship are more appropriate, so linear models are more reasonable.

As evident, the disadvantage of using linear models compared to nonparametric methods is the requirement for the linear mean function and homoskedastic variance. The relationship becomes even more complex as regressor terms are added.

Multiple Linear Models and Box-Cox Transformations

From the linear model (10) discussed previously, one may write,

$$\beta = \{\beta_0, \beta_1\}^T, \text{ and let } x_0 = \{1, 1, 1, \ldots, 1\}^T, \quad (16)$$

a column vector consisting of m "1's," where m is the number of observations. Then define the augmented matrix X as $X = \{x_0, x_1\}$ and denote all of the observations y_i as the column vector $\boldsymbol{y}$. The linear model may then be simply expressed as $y|X = X\beta + \varepsilon$, where X is the matrix of observations and β is the column vector of parameters. Thus expanding the number of predictor variables and associated coefficients transforms the simple linear model (10) into the multivariate model,

$$y|X = \beta_0 + \beta_1 x_1 + \ldots \beta_m x_m + \varepsilon \rightarrow y|X = X\beta + \varepsilon \quad (17)$$

$$\text{and } E(y \mid X) = X\beta \rightarrow \hat{E}(y \mid X) = X\hat{\beta}$$

with a single dependent variable, y, but multiple predictors referenced in the estimated parameter values in vector $\hat{\beta}$. The variance remains as in (12).

$$V(y|X) = V(X\beta + \varepsilon) = V(\varepsilon) = (\varepsilon^T \varepsilon) / n = \sigma^2 I, \text{ with } \hat{V} = \hat{\sigma}^2 I. \quad (18)$$

One notices that the variance is again constant (homoskedastic). Similarly, expanding the number of dependent variables results in the multiple linear model expressed as

$$Y|X = XB + E \text{ with } E(Y|X) = E(XB + \varepsilon) = XB \tag{19}$$

(a column vector), while $V(XB + E)=V(E)= \Sigma$, a symmetric covariance matrix. This matrix is diagonal if it is assumed that the coefficients β are independent of each other. Figures 10 and 11 illustrate the concept.

$$\begin{array}{ccccccc} y\,|\,X & = & \beta & + & X & *\beta & +\varepsilon \\ \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{bmatrix} & = & \begin{bmatrix} \beta_0 \\ \beta_0 \\ \beta_0 \\ \beta_0 \\ \beta_0 \end{bmatrix} & + & \begin{bmatrix} x_{11} \\ x_{12} \\ x_{13} \\ x_{14} \\ x_{15} \end{bmatrix} & *\begin{bmatrix}\beta_1\end{bmatrix} + & \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \end{bmatrix} \end{array} \quad \rightarrow \quad \begin{array}{ccccccc} y\,|\,X & = & X & * & \beta & + & \varepsilon \\ \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{bmatrix} & = & \begin{bmatrix} 1 & x_{11} \\ 1 & x_{12} \\ 1 & x_{13} \\ 1 & x_{14} \\ 1 & x_{15} \end{bmatrix} & * & \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} & + & \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \end{bmatrix} \end{array} \tag{20}$$

$V(y|X) = (\varepsilon^T \varepsilon) / n = \sigma^2 I$

Figure 10. Matrix Form of Simple Linear Regression Model.

Assume five observations of one dependent and one independent variable were taken. The simple model for each observation is expressed in matrix notation as follows. The advantage in using this matrix notation for simple models is that the application to models with multiple predictors is then straightforward.

$$Y \mid X \quad = \quad X \quad * \quad B \quad + \quad E$$

$$\begin{bmatrix} y_{11} & y_{21} \\ y_{12} & y_{22} \\ y_{13} & y_{23} \\ y_{14} & y_{24} \\ y_{15} & y_{25} \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{21} & x_{31} & x_{41} & x_{51} \\ 1 & x_{12} & x_{22} & x_{32} & x_{42} & x_{52} \\ 1 & x_{13} & x_{23} & x_{33} & x_{43} & x_{53} \\ 1 & x_{14} & x_{24} & x_{34} & x_{44} & x_{54} \\ 1 & x_{15} & x_{25} & x_{35} & x_{45} & x_{55} \end{bmatrix} * \begin{bmatrix} B_{10} & B_{20} \\ B_{11} & B_{21} \\ B_{12} & B_{22} \\ B_{13} & B_{23} \\ B_{14} & B_{24} \\ B_{15} & B_{25} \end{bmatrix} + \begin{bmatrix} E_{11} & E_{21} \\ E_{12} & E_{22} \\ E_{13} & E_{23} \\ E_{14} & E_{24} \\ E_{15} & E_{25} \end{bmatrix} \tag{21}$$

$$V(Y|X)=(E^TE) / n = \frac{1}{n} * \begin{bmatrix} E_{11} & E_{12} & E_{13} & E_{14} & E_{15} \\ E_{21} & E_{22} & E_{23} & E_{24} & E_{25} \end{bmatrix} \begin{bmatrix} E_{11} & E_{21} \\ E_{12} & E_{22} \\ E_{13} & E_{23} \\ E_{14} & E_{24} \\ E_{15} & E_{25} \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix} \tag{22}$$

Figure 11. Matrix form of Multiple Linear Regression Model.

Assume five observations of two dependent variables and two independent variables were taken. The linear model is expressed using the same notation as the previous example. The covariance matrix is diagonal if the error terms are considered to be independent, so that $\sigma_{12} = \sigma_{21} = 0$.

The multiple linear model provides some additional challenges. The assumptions associated with simple linear regression do not change; however, they are more difficult to investigate. Specifically, the assumption that the error is independently distributed with zero mean and homoskedastic variance requires that, for a single dependent variable,

$$E(y \mid X_i=x_i) = E(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n + \varepsilon_i) =$$

$$\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n, \text{ a hyperplane, and}$$

$$Var(y|X_i.=x_i)=\sigma^2 I. \tag{23}$$

Investigation of these assumptions is non-trivial. Investigation for multiple dependent variables is even more difficult.

Evidence of non-compliance is, however, to be investigated. It is often convenient to transform all pairs of variables so that their means are as linear as possible and to eliminate or combine highly collinear predictors. Multivariate normality ensures a linear mean function for any subgroup of two variables; however, the converse is not true. Even the transformations of the pairs of variables do not guarantee that the assumptions are met (Cook & Weisberg, 1999).

Consider adding an additional variable (supported population) to the previous example. A scatterplot matrix (Figure 12) suggests that there is significant heteroskedasticity between the population and dollar variables and population and workload variables. The 3-dimensional graph of the data against the $E(y|X)$ plane and associated residuals (Figure 13) reveals the heteroskedasticity as well.

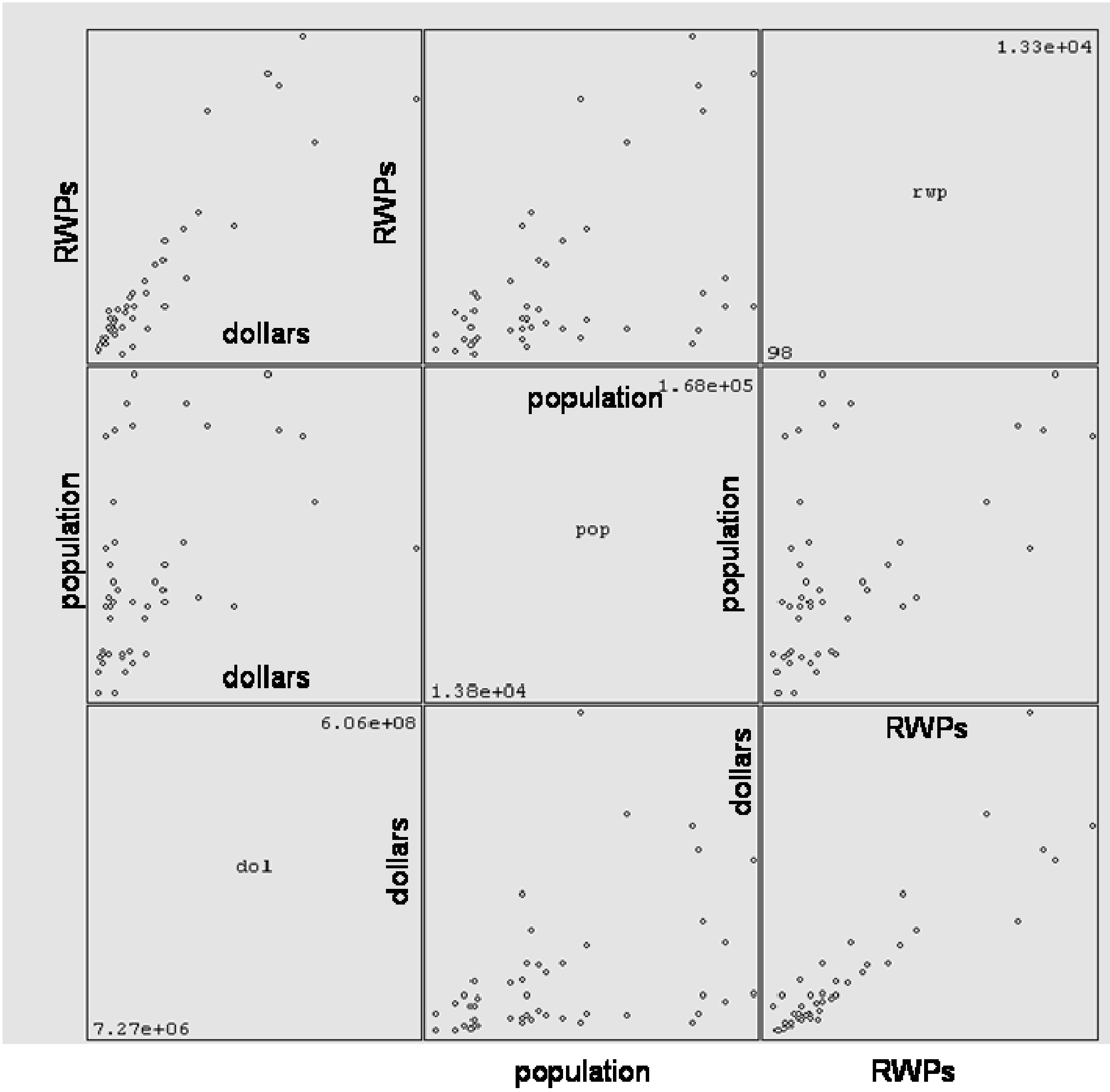


Figure 12. Scatterplot Matrix, Multiple Variables.

The scatterplot matrix above is organized as follows. In the top row, the vertical axis is *workload* ("rwp"), while the horizontal axes are *dollars* and *population* respectively. The second row has plots with *population* on the vertical axes, while *dollars* is plotted on the vertical axes of the third row. The spread of the plots associated with *population* suggests significant heteroskedasticity (as noted by the differential spread along the vertical axes), while the graph of *rwp* vs. *dollars* suggests increasing spread. Data transformations may improve the appropriateness of the multiple linear model.

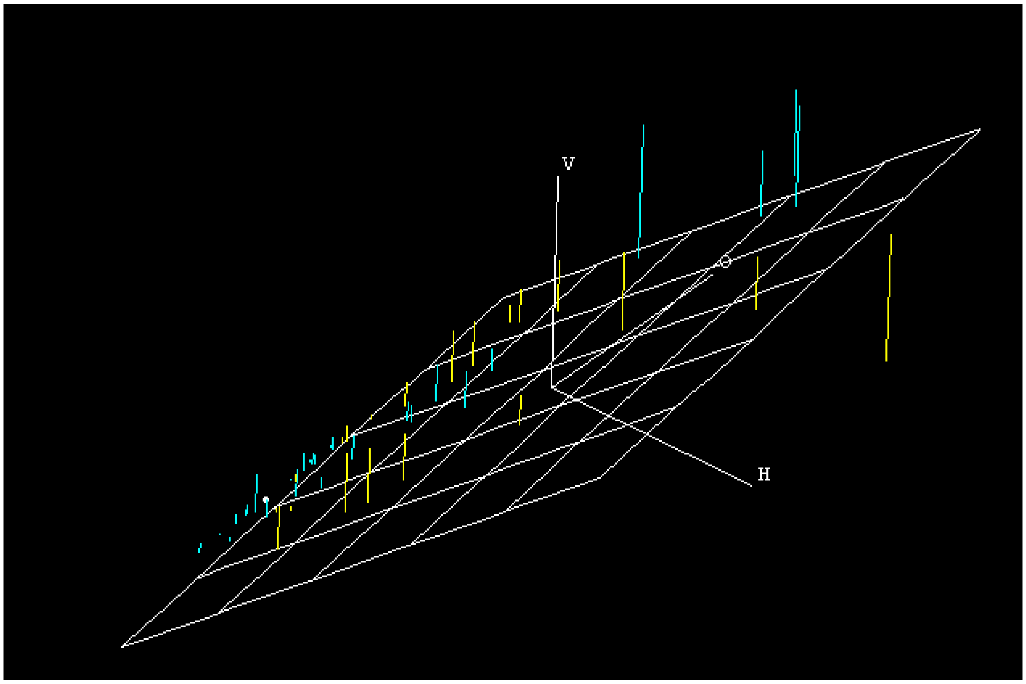


Figure 13. Three-Dimensional Plot of Heteroskedastic Residuals.

The 3-dimensional plot of the data points, the residuals, and the *E(y|X)* plane show clearly that the variance of the proposed model is not homoskedastic. As *population* ("O" axis, seen in the 2 p.m. position) and *dollars* ("H" axis, seen in the 4 p.m. position) increase, the residuals ("V" axis, seen in the 12 o'clock position) increase in magnitude. The assumptions required of the multiple linear model may be better met through transformation or the application of weighted least squares models.

Once again, transformations based upon the data should be investigated to improve model compliance with assumptions. Using the Box-Cox algorithm specified in (15), the optimal set of λ's are small for all variables (for *dollars*, λ=.15; for *workload*, λ=.27; for *population*, λ=.16.) These transformations improve the adherence to the assumptions as depicted in the next figures (Figures 14 and 15).

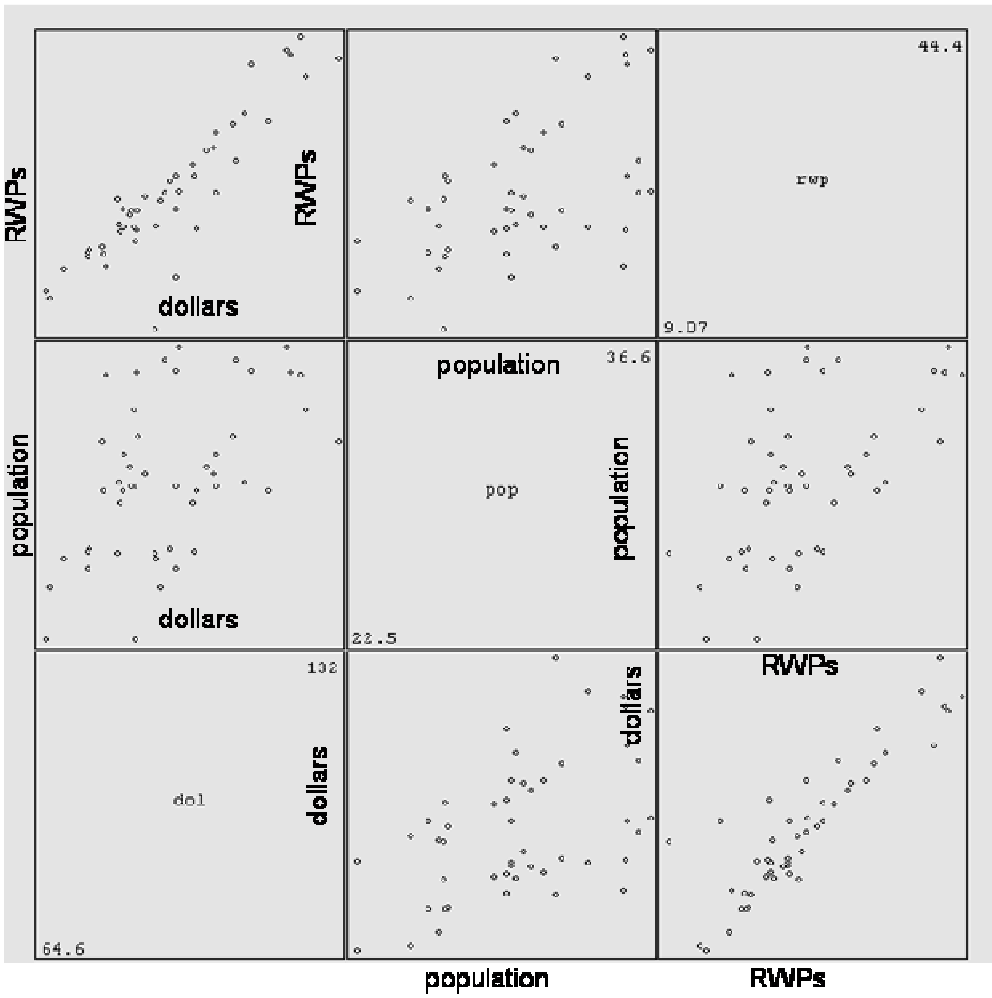


Figure 14. Scatterplot Matrix, Box-Cox Transformed Variables.

The scatterplot matrix of the transformed variables reveals that the relationship between *dollars* and *workload* is seemingly linear, while the relationship between *population* and both *dollars* and *workload* appear to be more linear than in the untransformed plots.

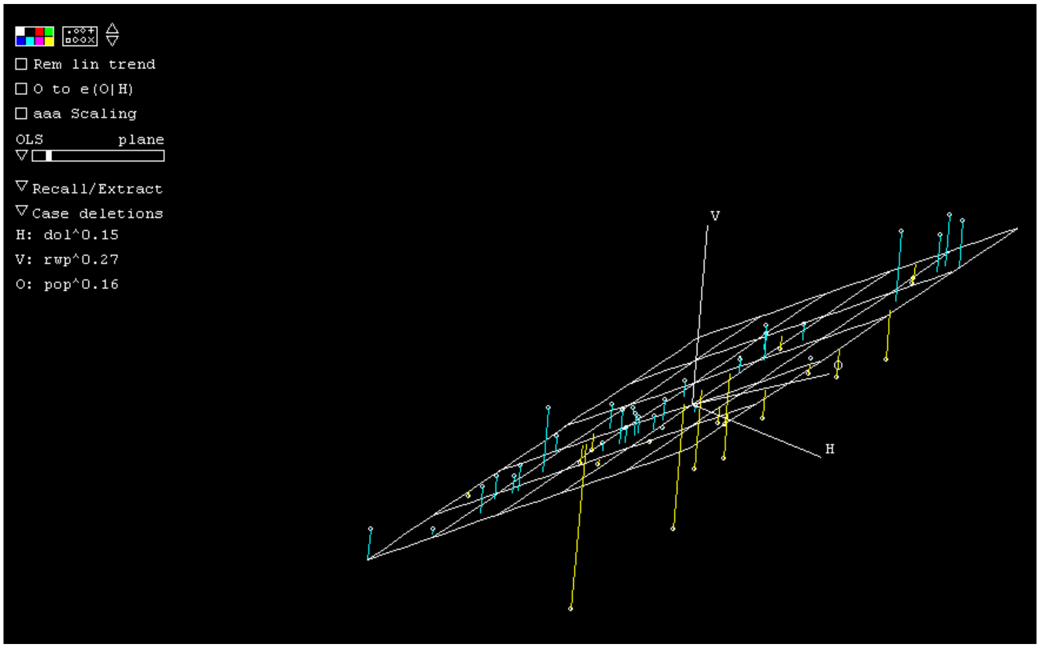


Figure 15. Three-Dimensional Plot of Residuals, Transformed Variables.
The plot of the transformed variable residuals against the expectation plane reveals an improved compliance with the multiple linear model assumptions. Again, the "V" axis shows the residuals, the "O" axis shows the population variable, and the "H" axis shows the workload variable.

The issue of independence of the observations *y* is also problematic. Specifically, time series data poses special concerns that must be handled accordingly. The Prais-Winsten (1954) transformation generates an autoregressive variable (AR(1)) by using the correlation between residuals and lagged residuals from a preliminary regression. The autoregressive variable is then included in the model. Metzger (2000) illustrates how least squares solutions may be adapted for dependent observations. This method will be discussed later.

Assuming that the multiple linear model is appropriate, it is simple to determine parameter estimates without making additional assumptions regarding the error distribution by minimizing the squared error (least squares solutions). Even if the assumptions do not hold, it may be possible to use a modification of the least squares method. Maximum Likelihood Estimation (MLE) of the parameters requires further distributional assumptions, although, for the normal distribution, the results are the same as the Ordinary Least Squares solutions.

Ordinary Least Squares Estimation (OLS)

The Least Squares approach for producing parameter estimates seeks to minimize the squared error. In 1812, LaPlace illustrated that the least squares solutions provided unbiased estimates regardless of the parent distribution. The Gauss-Markov Theorem as expanded by Aitken in 1934 provides that the least squares solutions are the best linear unbiased estimator for the parameters or any linear combination of parameters assuming that $\varepsilon \sim MVN(\vec{0}, \Sigma)$, where $\Sigma = \sigma^2 I$, finite, and independent of y. If the model is nonlinear and the errors are not normally distributed, then the least squares solutions usually do not provide the optimal result (Metzger, pg.150 and 160).

The OLS minimization problem is readily stated

$$\underset{\beta}{Min}\ (y - X\beta)^T(y - X\beta). \quad (24)$$

The matrix X and the vectors y and β are defined as previously. The gradient of the function is

$$\nabla = 2(X^T y - X^T X\beta) \propto X^T y - X^T X\beta\,. \quad (25)$$

Setting this vector equal to zero results in $\hat{\beta} = (X^T X)^{-1}(X^T y)$, which satisfies the first order conditions for maximization optimality. From inspection, it is clear that the Hessian matrix, the derivative of the gradient shown in (25), will be negative definite at the point $\hat{\beta} = (X^T X)^{-1}(X^T y)$ as the remaining components exist only on the diagonal and are negative constants. By definition, a matrix H is negative definite if $x^{\mathrm{T}}Hx < 0 \ \forall \ x \neq 0$ and such a matrix satisfies the necessary and sufficient conditions for maximization optimality (Nash and Sofer, 1996). The Hessian is expressed as

$$H = \begin{vmatrix} \frac{\partial^2 \ell}{\partial \beta_0^2} & \frac{\partial^2 \ell}{\partial \beta_0 \partial \beta_1} & \cdots & \frac{\partial^2 \ell}{\partial \beta_0 \partial \beta_n} \\ \frac{\partial^2 \ell}{\partial \beta_1 \partial \beta_0} & \frac{\partial^2 \ell}{\partial \beta_1^2} & \cdots & \frac{\partial^2 \ell}{\partial \beta_1 \partial \beta_n} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial^2 \ell}{\partial \beta_n \partial \beta_0} & \frac{\partial^2 \ell}{\partial \beta_n \partial \beta_1} & & \frac{\partial^2 \ell}{\partial \beta_n^2} \end{vmatrix} = \begin{vmatrix} -X^T X & 0 & \ldots & 0 \\ 0 & -X^T X & \ldots & 0 \\ \ldots & \ldots & \ldots & 0 \\ 0 & 0 & \ldots & -X^T X \end{vmatrix}.$$

When the cross-derivatives all are zero, selection of any non-zero x then ensures that the necessary and sufficient conditions, $x^{\mathrm{T}}Hx < 0$, are satisfied. Therefore, from (25), the estimate

$$\hat{\beta} = (X^T X)^{-1}(X^T y) \qquad (26)$$

provides the Ordinary Least Squares estimates of the coefficient parameters.

It is easy to show that these estimates are identical to the Maximum Likelihood Estimates under the assumption of normality. Specifically, one seeks to maximize the product of independent normal random variables. The likelihood function is simply

$$L(\mu \mid Y_1 \ldots Y_n) \propto \prod_{i=1}^{n} \exp\left(-\left(\frac{y_i - \mu}{\sigma} \right)^2 \right) \propto \sum_{i=1}^{n} -(y_i - \mu)^2 \approx \sum_{i=1}^{n} -(y_i - X\beta)^2. \quad (27)$$

One notices that maximizing (27) is equivalent to the minimization specified in (24).

Corrected Ordinary Least Squares (COLS) in Efficiency Analysis

One of the methods for evaluating efficiency involves Corrected Ordinary Least Squares (COLS). This method adjusts the OLS line upwards by the maximum of the residuals. Specifically, for the production frontier, the intercept is adjusted as follows

$$\hat{\beta}_0 * = \hat{\beta}_0 + \max_i(\hat{r}esidual_i) \text{ with } TE_i = \exp(-\hat{r}esidual_i + \max_i(\hat{r}esidual_i)) \quad (28)$$

where TE_i is the technical efficiency score for the *i*'th DMU (Kumbhakar and Lovell, 2000). Figure 16 illustrates the concept.

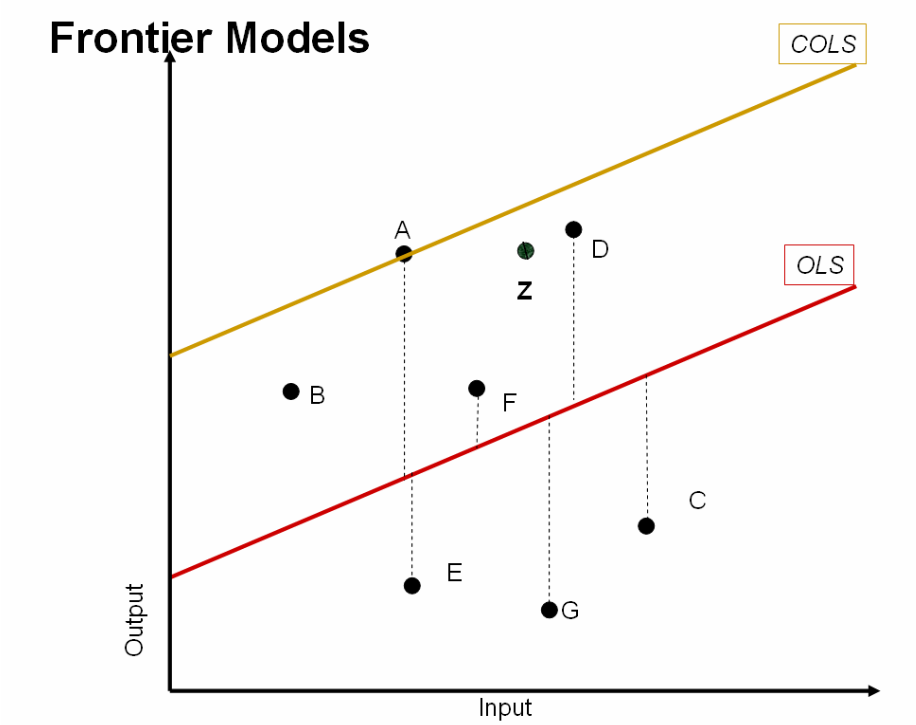


Figure 16. Corrected Ordinary Least Squares.

The Corrected Ordinary Least Square approach to efficiency shifts the OLS line by the maximum residual in the direction of the frontier. In the illustration above, one can see that the COLS line is parallel to the OLS line. In this way, all DMUs are bounded from above and the efficiency is measured as a distance function from the COLS line.

<u>Weighted Least Squares Estimation (WLS)</u>

Weighted Least Squares (WLS) allows the traditional linear model to account for non-constant variance. The WLS model is readily stated as:

$$y_i \mid X = X\beta + \frac{\varepsilon}{\sqrt{w_i}}, \text{ where } \varepsilon \sim N(0, \sigma^2). \qquad (29)$$

From (29), the expectation is simply $E(y_i \mid X) = E(X\beta + \frac{\varepsilon}{\sqrt{w_i}}) = X\beta$

and the variance is simply $V(y_i \mid X) = V(X\beta + \frac{\varepsilon}{\sqrt{w_i}}) = V(\frac{\varepsilon}{\sqrt{w_i}}) = \frac{\sigma^2}{w_i}$.

(Cook and Weisburg, 2000). Determining the weights is a function of the individual problem.

In a previous example, outpatient *workload* was calculated as a function of budgeted *dollars* for one specific year. Box-Cox transformations suggested that both variables might be transformed by use of the natural logarithm. Consider expanding this example and including two additional time periods. One might surmise that the variance of *log(workload) | log(dollars)* might not be constant over time. One can then select weights based upon the reciprocal of the sample variance or investigate Box-Cox transformations for all three years. In both cases, reduction of heteroskedasticity should occur, and the model should be more in compliance with (18). (Cook and Weisburg, 1999).

Feasible Generalized Least Squares (GLS and FGLS)

Least squares estimation is often not optimal for time dependent models such as multi-year studies. If one assumes that the value of a response variable in year *t* is correlated with the value of the response variable in year *t-1*, then the covariance matrix is non-diagonal. Given a non-diagonal error covariance matrix $\hat{V}$, the appropriate least squares solutions is no longer (26) but is expressed as follows:

$$\hat{\beta} = (X^T V^{-1} X)^{-1} (X^T V^{-1} y) \qquad (30)$$

Similar to (24), one seeks to solve the following minimization problem.

$$\underset{\beta}{Min}\ (y - X\beta)^T V^{-1}\ (y - X\beta). \qquad (31)$$

Normally, V^{-1} is not available, so feasible GLS is performed. Determination of $\hat{V}^{-1}$ from panel data may be done by estimating the V^{-1} matrix (Metzger, 2002). A technique for estimation of this matrix under the assumption that correlation is restricted to the last previous time period is known as the Prais-Winsten or AR(1) procedure.

The AR(1) assumptions are explained by analysis of a model. Consider the following function:

$$\begin{aligned} Cost_{it} &= \alpha + \vec{\beta}(X_{i,t}) + \lambda(X_{i,t-1}) + \varepsilon_t \qquad (32) \\ \varepsilon_t &= \rho_1 \varepsilon_{t-1} + v_t \\ v_t &\sim N(0, \sigma_v^2) \\ \sigma &= \frac{\sigma_v^2}{1-(\rho_1)^2} \end{aligned}$$

In (32), the correlation between time periods is based solely on the correlation between t and $t-1$. Specifically, the error covariance matrix (autocovariance) may be expressed as follows:

$$\sum = \sigma^2 \begin{bmatrix} 1 & \rho_1 & \rho_1^2 & \rho_1^3 \\ \rho_1 & 1 & \rho_1^3 & \rho_1^4 \\ \rho_1^2 & \rho_1^3 & 1 & \rho_1^4 \\ \rho_1^3 & \rho_1^4 & \rho_1^4 & 1 \end{bmatrix} \qquad (33)$$

It is simple to assume higher order auto-regressive models as well. (Fox, 2002).

The problem with using feasible Generalized Least Squares AR(1) with the number of observations less than the number of time periods is that the error covariance matrix is singular (Beck et al, 1993). It is therefore mathematically impossible to perform feasible GLS on the data under these conditions. Surprisingly (and as illustrated by Beck), many journal authors purport findings using feasible GLS when the number of time periods is less than the number of observations. In these cases (as is the case in this dissertation), Least Squares with Dummy Variables (LSDV) is often performed.

Least Squares with Dummy Variables (LSDV)

A simple solution to the feasible GLS problem is to assume fixed time and group effects and use the Least Squares Dummy Variable approach. This model is formulated readily as a fixed group and time effects model as follows (Greene, 2003).

$$Cost_{it} = [X_1 \dots X_n \ \alpha_1 \dots \alpha_{m-1} \ \gamma_1 \dots \gamma_{t-1} \dots \mu] \begin{bmatrix} \beta_1 \\ \dots \\ \beta_n \\ 1 \\ \dots \\ 1 \\ \dots \\ 1 \end{bmatrix} + \varepsilon_{it} \tag{34}$$

$$\text{subject to:}$$
$$\sum_t \alpha_i = \sum_t \gamma_t = 0$$
$$\text{where } \alpha \text{ are the between group effects,}$$
$$\gamma \text{ are withing group effects,}$$
$$\mu \text{ is the intercept}$$

These models allow for separate intercepts for each group.

Inferential Tests of Linear Models

Until now, the only distributional assumptions required for both the simple linear and multiple linear models was that the error term should have an expected value of zero and constant variance. It is often convenient to assume that variables are distributed normally: $y|X \sim N(X\beta, \sigma^2)$ and, for multiple dependent variables, $Y|X \sim N(XB, \Sigma)$, where $\Sigma = \sigma^2 I$. Hypothesis testing of models then becomes a comparison of sub model to full model Mean Square Error. This ratio of variance of the models forms the test statistics, and this ratio has an *F* distribution (assuming that the models are independent). Inferential analysis regarding individual coefficients normally involves the use of the generalized *Wald* statistic, which has approximately a *t*-distribution.

These statistics derive from the Likelihood Ratio Tests; however, their derivation is omitted.

<u>Collinearity</u>

Linear models are robust to small violations of homoskedasticity; however, collinearity of the regressor terms may pose greater threats, especially if one wishes to use the multiple linear model coefficient estimates in resource allocation models. In the most severe case (one or more regressor terms may be expressed as linear combinations of the others), the matrix $X^T X$ is not invertible, and so the estimator for β is not uniquely defined. As the matrix $X^T X$ approaches singularity, one can expect the coefficients to become increasingly unstable in direction and the standard error may be significant (Aczel, 1993). In this dissertation, the models used to recommend resource allocation must have stable coefficients, so handling collinearity is a necessity. In order to eliminate such problems, it is convenient to exercise restraint in variable selection (e.g., employ stepwise procedures or the use of the Mallow statistic for variable selection) and to employ data reduction techniques such as Principle Component Analysis (PCA) or Factor Analysis (FA).

Factor Analysis (FA) attempts to reduce multiple variables to a smaller set of factors. If the number of factors is chosen to equal the number of original variables and orthogonal directions are used, then the procedure is nothing more than Principle Components Analysis (PCA).

Principle Components Analysis in Linear Models

To illustrate PCA, assume that one believes the latent variable *Hospital Access* is associated with five different metrics m_i with mean μ_i and variance σ^2_i. Assume that these metrics are survey items, which are highly correlated. If the raw variables are included as predictors in a typical linear model for prediction of satisfaction (response variable y), the coefficients may be unstable and that the variance of the components may be large due to collinearity. Specifically, assume that the original linear mean function is improperly specified as in (10). To correct the collinearity, the mean function might be restated as:

$$E(y_1 \mid \lambda, m) = \beta_0 + \sum_{j=1}^{5} \beta_1 m_{1,j} \lambda_{j,1} + \ldots \sum_{j=1}^{5} \beta_5 m_{1,j} \lambda_{j,5} \qquad (35)$$

where the matrix Λ is chosen such that all vectors in $M\Lambda$ are orthogonal (zero covariance), k is less than or equal to the number of original variables (five in this case), and y and m variables are standardized by subtracting their expectation and dividing by their standard deviation. In matrix notation, the PCA linear model is simply $y|\lambda, M = M\Lambda\beta + \varepsilon$, where M is the matrix of all standardized observations m, Λ is a weighting matrix (Eigenvectors as will be shown), and β are the coefficients estimated as $((M\Lambda)^T(M\Lambda))^{-1}((M\Lambda)^T(y))$ as in multiple linear models (Mantha, 2004).

In the case of PCA, the matrix Λ is determined by maximizing the covariance in the following constrained optimization problem:

$$\underset{\Lambda}{Maximize}\, E(\Lambda^T (M^T M)\Lambda) \tag{36}$$

subject to $\Lambda^T \Lambda = I$, a normalizing constraint.

Forming the LaGrangian and solving the problem reduces to finding the Eigenstructure of the covariance matrix as illustrated.

$$\underset{\Lambda,\lambda^*}{Maximize}\, E(\Lambda^T (M^T M)\Lambda) + \lambda^* (\Lambda^T \Lambda - I)$$

$$\frac{\partial L}{\partial \Lambda} \propto (M^T M)\Lambda - \Lambda\lambda^* = 0 \rightarrow (M^T M - \lambda^* I)\Lambda = 0 \Rightarrow \left| M^T M - \lambda^* I \right| = 0 \tag{37}$$

The solution to the maximization problem is nothing more than the determination of the Eigenvalues from the characteristic equation. The largest Eigenvalue is associated with the Eigenvector that is the weight matrix for the first principle component. The second largest Eigenvalue produces the Eigenvector associated with the second principle component. The percent of variance accounted for by a specific Eigenvector (say E_1) may be calculated as $\sum_{i=1}^{n} \frac{E_1}{E_i}$. The matrix of Eigenvectors is often referred to as the loading matrix, as it determines the coefficient weights for use in transforming the original matrix, while the Eigenvalues are the variance components for the factors (Mantha, 2004).

Using PCA alone, however, retains a significant number of terms that may be less relevant. Factor Analysis (FA) considers elimination of terms through decision rules.

Factor Analysis

The linear factor analysis model for one latent variable may be expressed: $m|\ \Lambda, f = \Lambda f + \varepsilon$ with covariance matrix Σ, where m is one of the j data vectors (assumed to have been standardized), Λ is an unknown parameter matrix, f is the factor score vector, and ε is the independent error (noise) term. The rank of f is less than that of the rank of m, so that data reduction occurs. Factors are assumed to have a normal distribution and be independent of the errors. (Typically, factors may be construed to be independent themselves; however, the inclusion of an interaction term as a separate factor is sometimes used to ameliorate the requirement for such an assumption.) Errors are assumed to be independent and normally distributed. If these assumptions hold, then the correlation matrix ψ is a diagonal matrix. The assumptions are succinctly stated: $f \sim N(0,I)$; $\varepsilon \sim N(0, \psi)$ with ψ a diagonal matrix; f independent from ε. Of course, solving the original model is not feasible; however, since Σ is the covariance matrix, one may proceed as follows:

$$\Sigma = Cov(m)=Cov(\Lambda f + \varepsilon)=Cov(\Lambda f)+ \psi \text{ (by independence)}=$$

$$\Lambda Cov(f)\Lambda^T + \psi = \Lambda I \Lambda^T + \psi = \Lambda\Lambda^T + \psi . \qquad (38)$$

The relationships described by this covariance matrix are referred to as the factor analysis equations (Mantha, 2004). It is important to note that the matrix Λ is non-unique. One could find several equivalent formulations as readily illustrated.

$$m = \Lambda f + \varepsilon \rightarrow m = -\Lambda\theta^1 d + \Lambda\theta^1(\theta f + d) + \varepsilon \rightarrow m = \Lambda^* f^* + \varepsilon$$

(Yalcin & Amemiya, 2001). Therefore, oblique rotations might in some cases be considered along with the traditional orthogonal rotations. Thurstone (1947) provided five criteria for evaluating the loading matrix, which are useful for selection of rotations.

1. Each row should contain at least one zero.
2. For each column, there are at least as many zeros as there are columns.
3. For any pair of factors, there are some variables with zero loadings on one factor and large loadings on the other factor.
4. For any pair of factors, there is a sizable proportion of zero loadings.
5. For any pair of factors, there is only a small number of large loadings.

Solving the factor analysis equations is again a function of constrained optimization. Assuming that the variables are standardized, the covariance matrix Σ is nothing more than the correlation matrix. One seeks to find the solution to (25) for values of Λ and ψ. If Σ is of dimension $p \times p$ and Λ is of dimension $p \times m$ where $p>m$, then one must solve $(p / p+1) / 2$ equations (as Σ is symmetric) in $pm + p =$

$p(m+1)$ unknowns. If $m > (p-1)/2$, there exists no solution. Assuming that solution sets exist, one might choose the principal component set for inclusion, since they are orthogonal (e.g. maximize the variance capture of Λ.) The optimization problem is identical to that performed using PCA, except that one chooses the optimal subset by selecting the largest Eigenvalues associated with the Eigenvectors. A rule of thumb often employed is that Eigenvalues > 1 should be included in the equation, although Scree plots are used as well.

Unfortunately, the maximization problem does not result in a formal estimate for f, the factor loadings. If one intends to use the factor scores as replacements for the original variables, then the loadings are required. Bartlett's method is a simple solution to this problem. Once again, the procedure is equivalent to minimizing the squared error subject to a scaling matrix, similar to WLS (29) and feasible GLS (30)

$$\underset{f}{Min}\,(m-\Lambda f)\,\psi^{-1}(m-\Lambda f) \qquad (39)$$

which has the usual solution

$$(\Lambda^T \psi^{-1} \Lambda)^{-1}(\Lambda^T \psi^{-1} m). \qquad (40)$$

The calculus minimization is identical to that of least squares (Mantha, 2004).

The discussion of FA is limited to principal components, although other orthogonal methods exist (including Varimax, Quartimax, Transvarimax, Parsimax, and Equamax). For this dissertation, principal components are most useful.

Econometric Cost Functions

This dissertation relies on econometric models for analysis of cost functions in the resource allocation models. The Cobb-Douglas model is a multiplicative form for evaluating cost. It is expressed as

$$c = \beta_0 x_1^{\beta_1} x_2^{\beta_2} \ldots x_n^{\beta_n} \tag{41}$$

or in logarithm form as

$$\log(c) = \log(\beta_0) + \beta_1 \log(x_1) + \beta_2 \log(x_2) + \ldots \beta_n \log(x_n) \tag{42}$$

.

Thus the relation is loglinear. This formulation is nothing more than a linear in parameters model. The cost drivers are typically input prices and production output. The Cobb-Douglas assumes that the elasticity of substitution and scale are one, which may not hold (Creel, 2002).

Transcendental logarithmic (translog) models allow additional modeling flexibility. The cost function is not assumed to be linear and is modeled using a second order Taylor series expansion. For example, consider the following cost model derived from (41), $c = \beta_0 x_1^{\beta_1} x_2^{\beta_2}$. The translog formulation normally centers the regressor terms at zero by subtracting the means of the logarithms, e.g., ($z_1 = \log(x_1) - \log(\bar{x}_1)$). The translog approximation then becomes:

$$\log(c) = \log(\beta_0) + \beta_1 \log(z_1) + \beta_2 \log(z_2) + \frac{\beta_3 (\log(z_1))^2}{2} + \frac{\beta_4 (\log(z_2))^2}{2} + \beta_5 \log(z_1)\log(z_2). \tag{43}$$

Often, the translog formulation is coupled with appropriate Box-Cox transformations of the individual variables to produce the Generalized Translog (GT) formulation (Caves and Tretheway, 1980).

Any formulation of the hospital cost models requires assumptions pertaining to the nature of the cost structure, and it is interesting to note that the coefficient estimates for the translog model are not consistent. Specifically, the *plim* (limit in probability) for the estimates do not necessarily equal the parameters themselves unless the function belongs to the specified functional form (see Creel, 2004). It appears most useful to transform the data appropriately such that the functional relationships might be estimated consistently

Stochastic Frontier Analysis

Ferrier and Lovell (1990) illustrate that Stochastic Frontier Regression (SFA) and DEA may be used to crosscheck each other. This dissertation uses that approach. SFA is useful in that in provides either a production or cost frontier estimate along with inefficiency estimates.

Definitions

This dissertation analyzes cost frontiers for hospitals as part of the estimation of performance. A cost frontier may be defined as a function $c(y,w) = \min_x \{w^T x : x \in L(y)\}$, where *y* are the production outputs, *w* are the input prices for each input element *x*, and *L(y)* is the input isoquant with an input distance function equal to unity. In words, one seeks to find the optimal input isoquant that maintains outputs at the given level. The cost function is a concave, continuous function in *w* and semi-continuous in *y*, which satisfies the following additional properties (Kumbhakar and Lovell, 2000):

$$
\begin{aligned}
&c(0,w)=0; c(y,w)>0 \text{ for } y>0\\
&c(y,\lambda w)=\lambda c(y,w) \text{ for } \lambda>0\\
&c(y,w')\geq c(y,w) \text{ for } w'>w\\
&c(\lambda y,w)\leq c(y,w) \text{ for } 0\leq\lambda\leq 1
\end{aligned}
$$

The first property states that costs are zero when output is zero and costs are greater than zero for any non-zero amount of output. The second property states that cost increases directly (one to one) with input price increases. The final two properties are monotonicity properties.

SFA couples cost functions with efficiency. Specifically, this dissertation focuses on input-oriented measures of technical efficiency, which may be defined $TE_i(y,x)=\min\{\theta,\theta x\in L(y)\}$, where *L(y)* is the input isoquant and $0\leq\theta\leq 1$ is the measure of efficiency. Cost efficiency is also determined through SFA. Cost efficiency is the ratio of the minimum costs to observed costs and may be expressed

$$CE(y,x,w)=\frac{c(y,w)}{w^T x} \quad (44)$$

.From technical and cost efficiency, it is simple to derive a measure of input allocative efficiency, as it is simply the ratio of cost efficiency to technical efficiency,

$$AE(y,x,w)=\frac{CE(y,x,w)}{TE_I(y,x)} \quad (45)$$

(Kumbhakar and Lovell, 2000). These definitions lead to a discussion of SFA model formulation.

SFA Models

SFA is a relatively new procedure first developed in the 1977 works of Meeusen and van den Broeck (MB) and the work of Aigner, Lovell, and Schmidt (ALS) (Kumbhakar & Lovell, 2000). The traditional SFA model (Aigner, Lovell, and Schmit, 1977) assumes that the error term in the traditional regression model is actually two separate terms with different distributions.

SFA replaces the usual statistical regression formulation (17) in the following manner (formulated for a cost frontier),

$$y|X = f(X) + \varepsilon \rightarrow \; y|X = f(X) + \; v + u \qquad (46)$$
$$\text{where } v \geq 0$$

So the random term, ε, is replaced by two terms representing statistical error, v, and cost excesses, $u \geq 0$. Both are represented as statistical distributions. Normally, u is assumed to take on the distribution of an exponential, half-normal, or truncated normal to ensure that the non-negativity requirement is satisfied. Thus $y - u = f(x) + v$ where u shows that the cost would be reduced if the DMU under consideration were fully efficient.

This bifurcation of the error term is referred to as the “composed error” model, which means that the efficiency frontier, as determined by $y - u \leq y$ is stochastic (hence the name SFA). The random error term, *u*, therefore represents points both above and below the line.

This dissertation employs both the half-normal and truncated normal formulations. The basic cost equation is therefore specified in (46). Use of panel data simply requires the addition of an index associated with time so that (46) is written succinctly as

$$y_t|X_t = f(X_t) + u_t + v_t \qquad (47)$$
$$\text{where } v_t \geq 0$$

where the error vector and inefficiency vector are time dependent (index *t*). In this dissertation, (46) and (47) provide the model base with inefficiency modeled both as a half-normal and as a truncated normal.

Simple OLS provides a test for the presence of inefficiency. If the residuals are centered at zero, then there is little support for evaluating technical inefficiency. However, if the residuals are not centered at zero, then the error should be positively skewed (for a cost function) and testing for technical efficiency may be conducted (Kumbhakar and Lovell, 2000). A simple test for technical efficiency stems from Coelli (1995). Coelli proposed a hypothesis test of the skewness, illustrating that, asymptotically, the test statistic $\dfrac{m_3}{6m_2^3 / I}$ is distributed $N(0,1)$. Because this distribution is asymptotic, it may not work well for small samples. If negative skewness does exist and if it is statistically significant, then the next problem is estimating the parameter.

Estimation of Inefficiency in SFA Half-Normal Cost Models

If technical inefficiency is assumed to be distributed as a half-normal random variable independent from the random noise, then the distributions of the error terms for $\varepsilon = u + v$ are $v \sim N(0, \sigma^2_v)$, $u \sim N^+(0, \sigma^2_u)$. The distribution of u is twice the probability function of an unrestricted $N(0, \sigma^2_v)$ random variable by symmetry of the normal. The individual distributions for each random variable follow:

$$f(v) = \frac{1}{\sqrt{2\pi}\sigma_v} \exp\left(-\frac{v}{2\sigma_v^2}\right) \text{ and } f(u) = \frac{2}{\sqrt{2\pi}\sigma_u} \exp\left(-\frac{u}{2\sigma_u^2}\right) I_{[0,\infty)}(u). \tag{48}$$

From (48) and by independence of the random variables (a key assumption), the joint distribution is nothing more than

$$f(u,v) = \frac{2}{2\pi\sigma_u\sigma_v} \exp\left(-\frac{u^2}{2\sigma_u^2} - \frac{v^2}{2\sigma_v^2}\right) I_{[0,\infty)}(u) \tag{49}$$

.

From the joint distribution, the marginal distribution of the error is readily obtained by substituting for v ($\varepsilon = u + v$ so $v = \varepsilon - u$) and integrating out u. Specifically,

$$f(u,\varepsilon) = \frac{2}{2\pi\sigma_u\sigma_v}\exp\left(-\frac{u^2}{2\sigma_u^2} - \frac{(\varepsilon - u)^2}{2\sigma_v^2}\right) I_{[0,\infty)}(u)$$

$$f(\varepsilon) = \frac{2}{\sigma}\phi\left(\frac{\varepsilon}{\sigma}\right)\Phi\left(\frac{\varepsilon\lambda}{\sigma}\right),$$

$$\sigma = \sqrt{\sigma_u^2 + \sigma_v^2}$$

$$\lambda = \frac{\sigma_u}{\sigma_v}$$

$$\phi \equiv \text{pdf of standard normal}$$

$$\Phi \equiv \text{cdf of standard normal}$$

The distribution of technical efficiency is then calculated as

$$f(u \mid \varepsilon) = \frac{f(u,\varepsilon)}{f(\varepsilon)} = \frac{\dfrac{1}{\sqrt{2\pi}\sigma *}\exp\left(-\dfrac{(u-\mu^*)^2}{2\sigma^2 *}\right)}{1 - \Phi\left(-\dfrac{\mu *}{\sigma *}\right)} \qquad (50)$$

,

where $\mu^* = \varepsilon\sigma^2{}_u / \sigma^2$ and $\sigma^* = \sigma^2{}_u\sigma^2{}_v / \sigma^2$.

The distribution of technical efficiency given the entire error term is N^+ (μ^*, σ^2*). Jondrow, Lovell, Materov, and Schmidt (1982) disentangled the error term from the half normal noise term, deriving the following estimate for inefficiency.

$$E(u_i \mid \varepsilon_i) = \left(\frac{\sigma_u \sigma_v}{(\sigma_v^2 + \sigma_u^2)^{1/2}} \right) \left(\frac{\phi\left(\frac{\varepsilon_i * \sigma_u}{(\sigma_v^2 + \sigma_u^2)^{1/2} * \sigma_v} \right)}{1 - \Phi\left(- \frac{\varepsilon_i * \sigma_u}{(\sigma_v^2 + \sigma_u^2)^{1/2} * \sigma_v} \right)} + \left(\frac{\varepsilon_i * \sigma_u}{(\sigma_v^2 + \sigma_u^2)^{1/2} * \sigma_v} \right) \right) \quad (51)$$

$$\begin{gathered} \phi(a) \equiv \text{pdf of normal evaluated at } a \\ \Phi(a) \equiv \text{cdf of normal evaluated at } -\infty \text{ to } a \\ \varepsilon = u - v \equiv \text{joint error term} \\ u \sim N^+(0, \sigma_u^2) \equiv \text{technical inefficiency} \\ v \sim N(0, \sigma_v^2) \equiv \text{random noise} \end{gathered}$$

Since the distribution of technical efficiency is normal, either the expectation or the mode might be used to calculate technical inefficiency. The mode is also the MLE and is expressed as follows:

$$M(u_i \mid \varepsilon_i) = \left\{ \begin{matrix} \varepsilon_i \dfrac{\sigma_u^2}{\sigma^2} \text{ for } \varepsilon_i \le 0 \\ 0 \text{ otherwise} \end{matrix} \right\} \quad (52)$$

.

Estimates of cost efficiency are then calculated as $CE_i = exp(-\hat{u}_i)$ for either the expectation in (51) or the mode in (52). An alternate point estimator was provided by Battese and Coelli (1988). This estimator is expressed as follows.

$$CE_i = \left[\frac{1 - \Phi\left(\frac{\sigma * - \mu_i^*}{\sigma *} \right)}{1 - \Phi\left(\frac{-\mu_i^*}{\sigma *} \right)} \right] \exp\left(-\mu_i^* + \frac{1}{2} \sigma^{2*} \right) \quad (53)$$

Both estimators are inconsistent for cross-sectional data, because the variance of the conditional distribution is independent of the observation *i* (Kumbhakar & Lovell, 2000).

Efficiency Estimation in Truncated Normal SFA Cost Models

Similar to the half-normal model, the conditional distribution of the technical inefficiency given the entire error term is derived, and then the expectations are calculated for use in determining technical efficiency. Without derivation, the conditional distribution is expressed as in (50); however, μ^* and σ^* are expressed as $\mu^* = (-\sigma_u^2 \varepsilon_i + \mu \sigma_v^2) / \sigma^2$ and $\sigma_*^2 = \sigma_u^2 \sigma_v^2 / \sigma^2$. The estimators for inefficiency are similar to those provided in the half-normal model. The interested reader is directed to Kumbhakar and Lovell's discussion (2000).

Panel Data and SFA Cost Models

Analysis of panel data (multiple observations of the same DMU) provides a method for addressing three of the major limitations of SFA cross-sectional models. These limitations follow.

First, SFA relies on strong distributional assumptions, which may be relaxed by reliance on repeated observations. Second, fixed effects and random effects approaches using panel data alleviate the requirement for the cost inefficiency error component to be independent of the regressors. Third, consistent estimators for efficiency might be found if panel sizes are large enough, as the variance will tend towards zero.

It is simple to avoid distributional pitfalls by using the random effects model. In the random effects model, one makes the assumption that technical inefficiency is uncorrelated with the regressors and the random noise but may take any distribution with constant mean and variance. The noise is still assumed to have a mean of zero and constant variance. The rewritten model (where β_n are the coefficients associated with input prices and β_y are the coefficients associated with outputs) follows.

$$\log(Cost_{it}) = [\beta_0 - E(u_i)] + \beta_y \log(y_{it}) + \sum_n \beta_n w_{nit} + v_{it} + [u_i - E(u_i)] =$$

$$\begin{gathered}\beta_0 {}^* + \beta_y \log(y_{it}) + \sum_n \beta_n w_{nit} + v_{it} + u_i {}^* \\ \text{where } w_{nit} \text{ are input prices with } \sum_n \beta_n = 1, \\ y \text{ are the outputs,} \\ v \text{ is the random noise,} \\ \text{and } u \text{ is the random inefficiency}\end{gathered} \tag{54}$$

This model is often estimated using two-stage GLS methods. First OLS estimates of all parameters are generated along with estimates of the variance (using MLE or other methods). Second, the estimates of the variance are used to recalculate the parameter estimates using GLS (see Chapter 4). The estimates for cost efficiency (indexed by t time periods belonging to T) follow (Kumbhakar and Lovell, 2000).

$$\begin{gathered}\hat{u}_i {}^* = \frac{1}{T} \sum_t \log(Cost_{it}) - \hat{C}_{it} \\ \text{where } \hat{C}_{it} \text{ is the estimate for } \log(\mathrm{Cost}_{\mathrm{it}}) \\ \hat{u} = \hat{u}_i {}^* - \min_i \{\hat{u}_i {}^*\} \geq 0 \\ CE_i = \exp\{-\hat{u}_i\}\end{gathered}$$

The weakness in this formulation is that inefficiency is modeled as a constant throughout all time periods. Further, it is impossible to use GLS solutions (30) when the number of time periods is less than the number of observations as discussed earlier.

Cornwell, Schmidt, and Sickles (CSS) (1990) and Kumbhakar (1990) provided an approach, which allows for time varying efficiency. The model is formulated similar to (54), but it indexes the inefficiency term and provides an associated quadratic model to allow for variation.

$$
\begin{aligned}
&\log(Cost_{it}) = \alpha_{0t} + \sum_n \beta_n \log(x_{nit}) + v_{it} + u_{it} \rightarrow \\
&\alpha_{it} + \sum_n \beta_n \log(x_{nit}) + v_{it} \\
&\text{where } \alpha_{it} = \alpha_{0t} + u_{it} \\
&\text{and } \hat{\alpha}_{it} = \lambda_i + \mu_i t + \varsigma_i t^2
\end{aligned}
\tag{55}
$$

(Note: $\hat{\alpha}_{it}$ is an estimate of α_{it}.) In this model, the quadratic simplification of the intercept α_{it} allows efficiency to vary through time by producer. A simple way to generate estimates is to regress the response variable on Equation A with u_{it} removed to obtain estimates of the coefficients β_n. The residuals are then regressed on a constant, t, and t^2 to obtain estimates of λ, μ, and ζ for each DMU. To estimate the inefficiency term, it is simple to define $\hat{\alpha}_{0t} = \min\{\hat{\alpha}_{it}\}$ as the estimate for the intercept at time t. Then the estimate for cost efficiency is simply $\exp\{-\hat{u}_i\} = \exp\{\alpha_{0t} - \alpha_{it}\}$. These estimates of cost efficiency are consistent as the variance and bias converge to zero as number of observations and time periods go to infinity. It is also easy to estimate the model above using Maximum Likelihood Estimation if one makes assumptions regarding the distributions (Kumbhakar and Lovell, 2000.)

While the quadratic model shown in (55) is interesting, a simpler method was proposed by Lee and Schmidt (1993). Lee and Schmidt suggest that the inclusion of *T-1* dummy variables to measure time varying trends provides a reasonable solution. This approach is reminiscent of the Least Squares with Dummy Variables (LSDV) model (23), although MLE provides the estimates. The model specification they propose modifies (47) as follows.

$$Y_{it}|X_{it} = f(X_{it}) + u_{i1} * I_{\{t=1\}}\{0,1\} + u_{i2} * I_{\{t=2\}}\{0,1\} + \ldots u_{iT-1} * I_{\{t=T-1\}}\{0,1\} + v_{it} \quad (56)$$
$$\text{where } v_t \geq 0$$

The index i represents the DMU, while the index t represents the time period. The variables I are nothing more than dummy or indicator variables that are associated with the *T-1* time periods.

SFA Strengths and Weaknesses

SFA has been widely used in the healthcare environment (Dor, 1994); however, it has both strengths and weaknesses, which are delineated in Table 5. Notable weaknesses include the method's susceptibility to collinearity through misspecification and the requirement for monotonicity of relationships. A recent study of Dutch dairy farm environmental efficiency evaluated the results of estimates produced by SFA with those produced by DEA. This research revealed that, while the efficiency rankings are similar, the efficiency scores do vary. Furthermore, while DEA was able to produce estimates in all cases, SFA could not due to monotonicity and theoretical violations (Reinhard, Lovell, & Thissjen, 1997).

Table 5. SFA Strengths and Weaknesses.

Strengths and weaknesses of the SFA method are provided below.

Strengths	***Weaknesses***
Accounts for extreme data that may be attributed to inefficiency in other models	Sensitive to dependency of variables (collinearity), requiring careful selection of input / output variables
Error modeled with two parameters (truncated normal distribution)	Requires a priori selection of a model

Bayesian Models

Bayesian theory derives from the posthumous work of Reverend Thomas Bayes. His work was written in the 1750s and published in 1763. In contrast to empirical or frequency interpretation (e.g., von Mises), one notes that Bayesian approaches measure probabilities associated with belief. LaPlace formalized Baye's work by providing the next equation, which he applied to celestial mechanics, medicine, and jurisprudence

*P (A B) = P (B | A) * P (A)* (Metzger, 2002).

It is easy to show that the following relationships hold (discrete case shown).

$$P(A \mid B) = \frac{P(AB)}{P(A)} = \frac{P(A \mid B)P(B)}{P(A)} \; and \qquad (57)$$

$$P(A) = \sum_i P(A \mid B_i)P(B_i) \text{ so}$$

$$P(A \mid B) = \frac{P(A \mid B)P(B)}{\sum_i P(A \mid B_i)P(B_i)}$$

From this equation, there exists a logical connection to theory testing. If one defines *P(A)* as the prior belief probability that a hypothesis is true and *P(B)* as the probability that the observed results were obtained, then the resulting equation (the posterior) is a probability measure of the accuracy of the beliefs.

. Upon closer analysis of this simple equation, one notes that the calculation of probabilities based upon the prior beliefs may be done iteratively by using the posterior probability as the prior in the next sample or experiment.

The following equation illustrates the relationship of Baye's probability statement to density calculations. The conditional Probability Density Function (pdf) of a series of observations $(x_1, x_2 \ldots x_n) \in X$ with a single parameter μ is expressed $f(x \mid \mu)$. In the case of discrete variables, the Probability Mass Function (PMF) may be expressed similarly or as $P(x \mid \mu)$. The distribution of one's belief about the parameter is expressed $f(\mu)$. The unconditional distribution of the random variable X is $f(x)$. These expressions lead to the below equation, which stems directly from (57).

$$\pi(\mu \mid x) = f_{posterior}(\mu \mid x) = \frac{f(x \mid \mu) * f_{prior}(\mu)}{f(x)} \propto f(x \mid \mu) * f_{prior}(\mu) \tag{58}$$

Updating the priors by current information is an advantage of the Bayesian methodology. The equation above assumes priors only on the parameter μ; however, this is a simplification for illustrative purposes. The Bayesian hierarchical posterior follows. Assume that one has a series of observations $(x_1, x_2, \ldots x_n) \in X$ and a series of parameters $(\mu_1, \mu_2, \ldots \mu_n) \in \mu$, $(\theta_1, \theta_2, \ldots \theta_n) \in \theta$, where θ is a parameter of the probability distribution for μ. These expressions lead to the following derivation, which expands (58)

$$\pi(\mu,\theta \mid x) = \frac{f(x \mid \mu) f_{prior}(\mu \mid \theta) f_{prior}(\theta)}{f(x)} \propto f(x \mid \mu) f_{prior}(\mu \mid \theta) f_{prior}(\theta) \tag{59}$$

It is clear from the previous equation that series of nested parameters may not simplify readily. Determining the likelihood function for the parameters becomes quite problematic. Empirical methods for handling these nested distributions via simulation are important and are included in Appendix 5.

Synopsis

In this chapter, the foundation for the empirical analysis has been provided. Efficiency analyses methods, linear models, Stochastic Frontier Analysis, and Bayesian estimation techniques were discussed in detail. This chapter provides the basis for all the empirical work in the remainder of the dissertation. With this discussion complete, one may turn to the efficiency analyses.

Chapter 4: Data Sources, Models, Descriptive Statistics, and Transformations

Data Sources

The military healthcare system relies on a collection of computer systems that do not all communicate adequately with each other. A single data repository (the M2 Database) has partially addressed this problem. The primary sources of data for this study include the Medical Expense and Performance Reporting System (MEPRS), the M2 Database, TRICARE Management Agency (TMA) Workload Reports, and the TRICARE Beneficiary Survey data (2001-2003). These sources are similar and in many cases identical to the data sources used in the *2002 TRICARE Report to Congress* and West's (2000) study of the Army Medical Department cost structure. A cursory description of the databases follows.

A MEPRS User's Guide (2002) provides a succinct description of this accounting database system: "The purpose of the MEPRS for DoD Medical Operations is to provide consistent principles, standards, policies, definitions, and requirements for accounting and reporting of expense, manpower, and performance by DoD fixed military medical facilities". Using these specific purposes, MEPRS provides in detail: uniform performance indicators, common expense classification by work center, uniform reporting of personnel utilization data by work center, and a cost assignment methodology." While the quality of data has sometimes been questioned, TMA and senior AMEDD leaders use the information for decision support, and the data quality has improved accordingly.

The M2 Database is the repository for multiple medical computer systems, including sources such as the Composite Health Care System (CHCS), the Ambulatory Data System (ADS), and others. Standardized inpatient and outpatient workload are available through the M2. The workload is standardized by weighting the visit or inpatient stay based on the associated acuity. Standardized inpatient workload is known as Relative Weighted Product (RWP), while standardized outpatient workload is known as Relative Value Units (RVU).

A clear definition of RWP derives from the *MHSS Workload Primer*. "The RWP is a weighted workload measure reflecting case complexity and length of stay. The general process for calculating RWP is based on the patient's source of admission (direct, transfer, or birth), length of stay, and disposition status (normal, transfer, discharged to nursing facility, etc.) in conjunction with the Diagnosis Related Group (DRG) weights and outlier thresholds" (Coventry, Gromadzki, Kiernan, Rogers, Smith, Spivey; 1995). For example, a patient receives a craniotomy (DRG 001) and stays in the hospital 10 days. The base RWP for this hospital transaction is 3.6537. If the patient stays longer, the RWP increases. For comparison, a rectal resection (DRG 146) with the same length of stay is weighted as 2.6146 (base weight).

For outpatient services, Relative Value Units (RVU) are a measure of the resources associated with a procedure (Current Procedural Terminology or CPT) and are designed to reflect physician work, practice expense, and physician liability. Medicare transitioned to the use of Resource-Based Relative Value System in the early

1990s, and the military has since adopted it. As an example, the initial visit for a normal newborn (Procedural Code 99431) was weighted as 1.17 in 2005.

For analysis of purchased care, the available data source was the TRICARE Management Agency (TMA) Workload Reports (see http://199.211.83.250/Reports/ for examples). These reports summarize both cost and utilization information for 26 medical specialties (inpatient and outpatient care). These reports provide general information (albeit imperfect) regarding the civilian network costs. Unfortunately, weighted workload information was unavailable for some regions; however, this problem was handled conservatively in favor of the military direct care facilities as will be discussed in Appendix 8. (Weights of one were assigned to all workload generated by civilian facilities.)

The *2002 TRICARE Report to Congress* used the Health Care Survey of Department of Defense Beneficiaries to Benchmark both Quality and Access within the system. This dissertation employs the same technique. "To fulfill 1993 National Defense Authorization Act requirements, the Health Care Survey of Department of Defense Beneficiaries (HCSDB) was developed by the TRICARE Management Activity. HCSDB questions on satisfaction with and access to health care have been closely modeled on the Consumer Assessment of Health Plans Survey (CAHPS®) program. CAHPS® is a nationally recognized set of standardized questions and reporting formats that has been used to collect and report meaningful, reliable information about the health care experiences of consumers. Because the HCSDB uses

CAHPS® questions, TRICARE (DoD's health plan) can be benchmarked to civilian managed care health plans (HCSDB, http://www.tricare.osd.mil/survey/HCsurvey/.) The HCSDB is stratified by three groups: enrollment type, beneficiary type, and geography. The survey is sent to 45,000 individuals and received a response rate of about 31% in 2003. Responses were weighted to account for non-response bias, e.g., typically $w(h,i) = \frac{N(h)}{n(h)}$, where *w(h,i)* is the weight of the ith observation in the *h*th stratum, *N(h)* is the number of total beneficiaries in the *h*th stratum, and *n(h)* is the number of sampled beneficiaries in the *h*th stratum (*Adult Technical Manual, 2003 HCSDB*).

With the primary data sources specified, models for determining efficiency and allocating resources must be developed.

Efficiency Models

For this dissertation, BCC-I and RAM models are the primary tools for evaluating efficiency. The variables selected for use in all the models reflect the managerial emphasis of the military healthcare system: workload, cost, quality and prevention, access, readiness requirements, and training.

<u>Output Selection</u>

Military healthcare DEA studies such as Charnes et al (1985) and Coppola (2003) to date have focused on the traditional aspects of facility production. Charnes' seminal work selected Relative Weighted Product (RWPs) for measuring inpatient workload, clinic visits for measuring outpatient workload, and personnel trained excluding interns and residents. These workload measures are appropriate with some modernization.

In this dissertation, RWPs are also selected as the measure of inpatient workload, while Relative Value Units (RVUs) serve as the standardized measure of outpatient workload. (RVUs were not available in 1985). In the civilian network, access to RWPs and RVUs is currently problematic. A recent analysis done by the Center for AMEDD Strategic Studies indicates that the inpatient and outpatient weights for civilian network facilities in Army Region 6 are universally greater than one and universally greater than those of the military facilities. As a conservative approach and in absence of concrete weights for all regions, weights of one were assigned to civilian workload. This approach probably favors the military facilities when cross-comparisons are made, as the civilian network generally has higher average acuity (seriousness of the average condition) for patient visits (personal communication, Dr. Barbara Wojcik, Center for AMEDD Strategic Studies). When available, weights associated with the civilian workload will be applied.

The approach to measuring training differs somewhat from previous studies. Graduate Medical Education (GME) is chosen as the training metric of interest instead of personnel trained, as it is the largest (in terms of dollars) training program within the military facilities. In order to account for differences in structure associated with training, facilities with large GME programs (as determined by the percent of adjusted dollars spent on GME) were analyzed separately from facilities with no GME program or small GME programs (see following figure). An analysis of the inflation-adjusted ratios of GME expenditures to total expenditures by year is shown in Figure 17 below. This shows that (almost) all MEDCENS have ratios greater than 1%, while community hospitals have ratios below 1% As a conservative measure, GME costs were deducted from total yearly costs for the facilities with GME programs (the MEDCENS). The GME programs should be subjected to separate analysis based on projected needs of the military. GME costs in the civilian network are invisible to the payer.

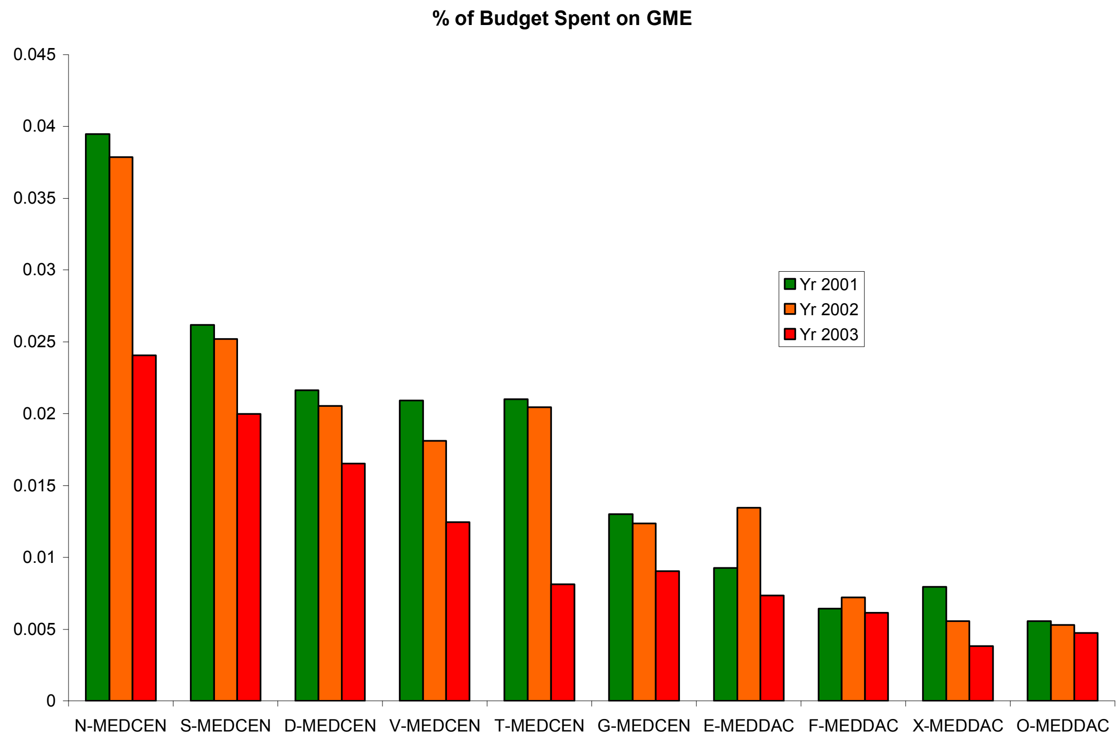


Figure 17. GME Expenditure Ratios.

The ratio of dollars coded "FAM" in MEPRS (roughly the GME expenditures) versus the total dollars expended suggests that there exists a reasonable bifurcation at .01. One might note that the ratio of recorded expenditures has declined from 2001 through 2003.

Coppola (2003) selected outpatient visits, Case Mix Adjusted Discharges (CMAD), RWPs, and live births as output metrics. Outpatient visits are unstandardized and less useful than RVUs, which are used herein. Specifically, a visit to the emergency room for a headache is less intensive than a visit to the emergency room for cardiomyopathy. While visits reflect output, they do not measure it as accurately as RVU's. Standardized workload metrics are required. For this dissertation, Relative Value Units are used to measure outpatient workload. Since RVUs are nothing more than outpatient visits weighted by a measure of acuity, both complexity and acuity for the outpatient visits are reflected in these numbers.

Coppola selects both RWPs and CMAD as output variables. He defines CMAD as the *Case Mix Index* * Σ*Discharges*. The *MHSS Workload Primer* defines

$$\text{Case Mix Index (CMI) as CMI} = \frac{\sum RWPs}{\sum Dispositions}$$

http://www.tricare.osd.mil/tma/hpae/primword.html). Dispositions should differ very little from discharges (possibly only by the number of deaths). With this definition, CMAD is nothing more than

$$\text{CMAD} = \frac{\sum RWPs}{\sum Dispositions} * \Sigma Discharges \approx \sum RWPs\,.$$

Hence RWP's and CMAD are nearly identical, and only one should be used. The live birth metric is also redundant as it is already a component of RWPs.

Other output metrics were considered for inclusion, and, in fact, some were originally evaluated in preliminary analysis. Similar to the Charnes et al 1985 study, occupied bed days and dispositions were evaluated for inclusion; however, these outputs provide the same information as the RVUs and RWPs, as illustrated in the next correlation matrix (Table 6). Because of the richness of the RVU and RWP metrics, no other additional workload metrics are required to capture both the complexity and the volume of workload.

Table 6. Correlations for Several Considered Output Metrics.

The inclusion of bed days and dispositions would provide redundant information. It should be noted that, while the collinearity is perfectly acceptable for DEA models, other models prefer nearly independent outputs. The collinearity of RVUs and RWPs will be addressed later. Asterisked items are statistically significant below $p < .001$. Note: The correlations were calculated using the n=72 dataset.

Correlations for Considered Metrics

		A.	B.	C.	D.
A.	Dispositions(000s)	1			
B.	Occupied Bed Days (000s)	.912**	1		
C.	RVU (000s)	.916**	.898**	1	
D.	RWP (000s)	.890**	.987**	.878**	1

Absent from the above studies are metrics for readiness, quality / prevention proxies, and access to care proxies. Readiness in this research is defined as a critical, non-discretionary output required of hospitals and is measured in terms of facility cost (MEPRS code G). Specifically, the largest part of the readiness mission entails deployment of personnel for real-world contingencies or required field training exercises and administration and planning associated with deployments. This cost is subtracted directly from the dollars of input to the facility. Unfortunately, the full cost of readiness is not measured in dollars. There are clear secondary and tertiary effects on quality, access, and other metrics as readiness requirements as a percent of budget increase. An illustration of this effect is provided in Figure 18.

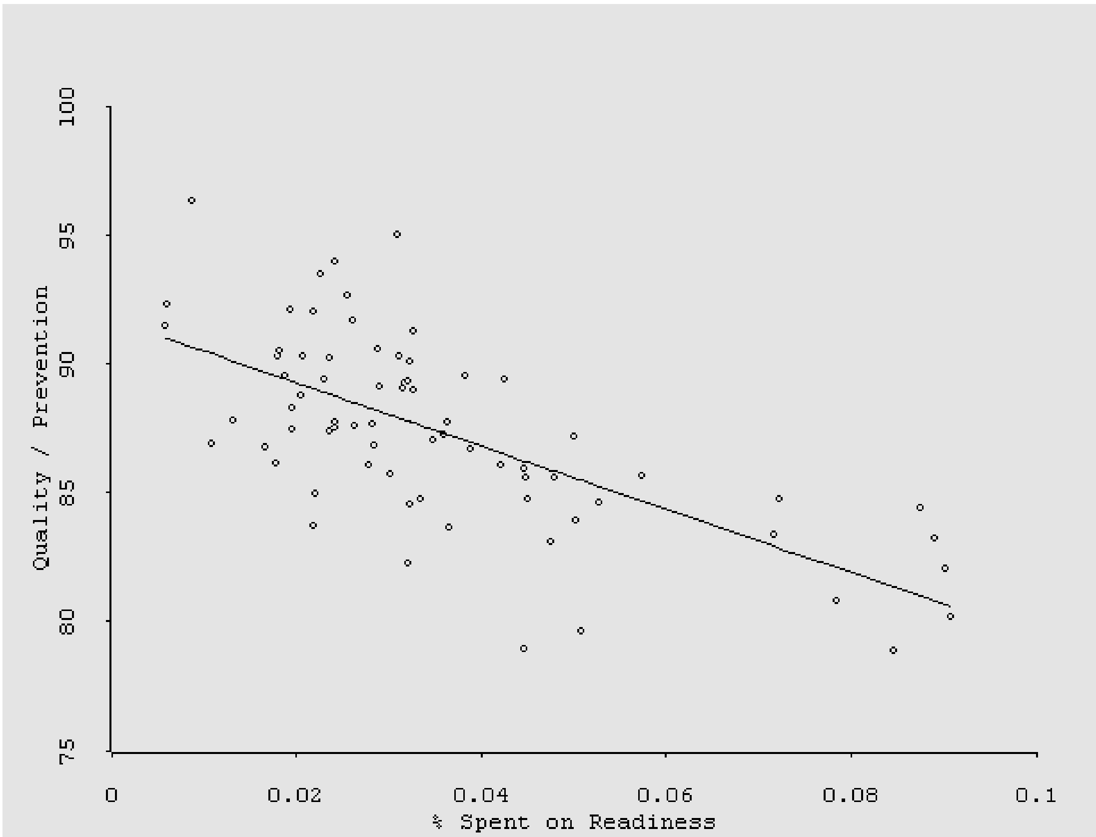


Figure 18. Secondary Effects of Readiness Expenditures.
Increases in readiness expenditures as a percent of budget have effects that cannot be isolated by simply removing the amount of expenditure from total cost. In this graph, one can readily see that the quality / prevention composite metric decreases as the percent spent on readiness increases.

Quality of care and prevention are proxied by a composite metric, which incorporates blood pressure readings, pap smears, mammography, prenatal care, and cholesterol screening (see Table 7). While this metric is certainly not comprehensive, it does measure items which relate to both prevention and quality. The inclusion of this metric is fully congruent with the *2002 TRICARE Report to Congress*. The preventive care composite score is generated as a percent of the care that all recipients should have and actually did receive. For example, if out of 100 required mammographies, 90 were performed the prevention score for that facility is 0.9. The exact metrics that generate the composite score are listed in the subsequent table. Conveniently, the survey is stratified based on TRICARE categorization. While there is no clean division for comparing the civilian network components to the military components (e.g., a TRICARE Standard patient might seek care in the military facility), there is some merit in assigning the PRIME respondents scores to the military facility and the other respondents scores to the civilian network with the understanding that there exists measurement error.

Table 7. Quality / Prevention Score Calculation.

Calculation of the prevention composite score is based on these questions from the HCSDB survey. Source: HCSDB survey.

When did you last have a blood pressure reading?

- Less than 12 months ago
- 1 to 2 years ago
- More than 2 years ago

Do you know if your blood pressure is too high or not?

- Yes, it is too high
- No, it is not too high
- Don't Know

When did you last have a Pap smear test?

- Within the last 12 months
- 1 to 3 years ago
- More than 3 but less than 5 years ago
- 5 or more years ago
- Never had a Pap smear test

Are you under age 40?

- Yes
- No

When was the last time your breasts were checked by mammography?

- Within the last 12 months
- 1 to 2 years ago
- 2 years to less than 5 years ago
- 5 or more years ago
- Never had a mammogram

In what trimester is your pregnancy?

- First trimester
- Second trimester
- Third trimester

In which trimester did you first receive prenatal care?

- First trimester
- Second trimester
- Third trimester

When did you last have a cholesterol screening, that is, a test to determine the level of cholesterol in your blood?

- Less than 12 months ago
- 1 to 2 years ago
- More than 2 but less than 5 years ago
- 5 or more years ago
- Never had a cholesterol screening

Access to care is generated from an HCSDB composite survey item, *Getting Needed Care*. (*Getting Care Quickly* was considered for inclusion as well; however, the high correlation between the two survey measures, .816, rendered the additional measure redundant.) For *Getting Needed Care*, the HCSDB composite score was generated by calculating the percentage of responses in the "Not a problem" category based on the survey items in Table 8.

Table 8. Access Score Calculation.

Getting Needed Care — Composite

With the choices your health plan gave you, how much of a problem, if any, was it to get a personal doctor or nurse you are happy with? (dangling preposition in original)

A big problem
A small problem
Not a problem

In the last 12 months, how much of a problem, if any, was it to get a referral to a specialist that you needed to see?

A big problem
A small problem
Not a problem

In the last 12 months, how much of a problem, if any, was it to get the care you or a doctor believed necessary?

A big problem
A small problem
Not a problem

In the last 12 months, how much of a problem, if any, were delays in health care while you waited for approval from your health plan?

A big problem
A small problem
Not a problem

A full display of the output variables is provided in Table 9. With the output variables defined, the input variables require specification. A review of input variables used in previous studies is appropriate as well.

Input Selection

Charnes et al (1985) selected multiple inputs including Full Time Equivalents (FTEs) by type and dollars budgeted for inpatient and outpatient care. Similarly, this dissertation captures FTEs in the computation of dollars budgeted for the facility. Inpatient and outpatient dollars are aggregated. [Note: While disaggregation is possible, there is loss of accuracy and clarity in doing so, as the stepdown accounting system used by facilities involves semi-arbitrary cost assignments. Further research would disaggregate the costs by MEPRS code A (inpatient) and code B (outpatient) services.]

Military facility dollars were included and calculated using a multi-stage process, which generated an information-rich metric. First, the MEPRS total annual expense values were used as the base. Second, readiness (MEPRS G coded expenditures), pharmacy expenses, and GME expenses (if any) were deducted to account for nondiscretionary variables and to apply different inflation rates to pharmaceuticals versus other medical expenditures. Next, 2001 and 2002 data were converted into 2003 dollars based on the Bureau of Labor Statistics (http://stats.bls.gov/cpi/) data for cost of medical care and cost of medical supplies and prescription drugs, similar to the adjustments applied in the *2002 TRICARE Report to Congress*. (These inflation adjustments were also applied to the civilian network dollars as well.) Finally, a 3.5% Major Construction of the Army (MCA) adjustment was applied identically as in the *2002 TRICARE Report to Congress* to account for facility maintenance cost that is not part of the facility budget. The overall adjusted budgeted dollars is expressed simply as:

$$Adj\$_{i,t} = ((TC\$_{i,t} - RX\$_{i,t} - GME\$_{i,t} - Rdy\$_{i,t}) * (1 + Inf1_t) + RX\$_{i,t} * (1 + Inf2_t)) * 1.035 \quad (60)$$

where i is the facility index, t is the time index (2001, 2002, or 2003), *Adj$* is the adjusted dollars for analysis, *TC$* is the total unadjusted dollars, *RX$* is the pharmacy expenditures, *Rdy$* are the expenditures on readiness, *Inf1* is the medical inflation rate, and *Inf2* is the pharmacy inflation rate specified by:

$$Inf1 = \begin{cases} 0.089076 & \text{if } t = 2001 \\ 0.040266 & \text{if } t = 2002 \\ 0 & \text{if } t = 2003 \end{cases}, \quad Inf2 = \begin{cases} 0.097892 & \text{if } t = 2001 \\ 0.059655 & \text{if } t = 2002 \\ 0 & \text{if } t = 2003 \end{cases}.$$

Coppola (2003) selected as input variables total number of operational beds in the hospital, total FTEs, total number of services provided, and total dollars budgeted (unadjusted). The number of operational beds may be a measure of size (although there are many different capacities reported in the military), but (unfortunately) the definition is not standardized within the military framework. Further, bed capacity is a function of staffing (e.g., the availability of skilled nursing staff determines the existence of bed capacity) and not all operational beds are alike (e.g., Cardiac Care Ward versus Intermediate Care Ward beds). A more stable variable for indicating bed capacity is FTEs, which Coppola includes. MEPRS includes FTE expenditures in its detailed cost reports, so, for overall efficiency determination, it is important to consider FTEs a subcomponent of cost in order to prevent double counting system inputs.

The total number of services provided is an interesting variable that addresses complexity. In this dissertation, complexity is accounted for by the selection of RWPs and RVUs as services are not equally intense (e.g., the presence of an oncology clinic is likely to be more resource intensive than the presence of a dermatology clinic).

The primary inputs for the DEA analysis in this dissertation are a non-discretionary population input (which includes both enrollment and population) and the adjusted budget. However, in Chapter 6 it is shown that replacing the composite enrollment/population input with enrollment alone produces no change in the DEA model results. Cost of FTEs is included in the budget figures. The population inputs (both enrollment and raw population) derive from the M2 database and reflect potential demand. There is additional demand that remains uncaptured. Rotational units (e.g., units from Fort Bragg visiting Fort Polk for training) and persons seeking treatment out of their geographic area are not captured. Further research might include a relatively new metric known as "reliant users." This metric would attempt to estimate the expected number of persons actually using the facility.

Synopsis of Variable Selection for the DEA Analysis

The analysis of input and output variables has resulted in a parsimonious set that captures cost, quality, access, training, readiness, and workload information. Further, the variable set addresses complexity across facilities by using weighted work products, readiness and training differentials by providing adjustments to the budgeted dollars, and structural differentials by capturing population information. These variables and their sources are summarized in the following table. Normalizing transformations, while not required for the DEA models, are definitely required for the SFA and linear models. These transformations are also used in the efficiency analyses for consistency and are provided later in this chapter.

Table 9. **DEA Input and Output Measures.**

The table reflects the parsimonious set of variables (prior to transformation) selected for analyzing facility performance.

Variable and Input / Output Status	**Data Source, MTF**	**Data Source, Network**
Cost Measure (Input): Expenditures less GME and Readiness Costs and inflated in two parts to 2003 dollars	MEPRS	CHAMPUS Workload, TMA
Population Measures (Non-Discretionary Input): Population and Enrollment	M2 Database	M2 (Population less Prime)
Quality Measure (Output): Prevention (BP, mammo, 5 quality measures)	DoD Survey	DoD Survey
Access / Satisfaction Measure(Output): Getting Needed Care metric from TRICARE PRIME survey	DoD Survey, Prime Members	DoD Survey, Standard / Extra Members
Workload (Output): Relative Value Units (weighted outpatient workload) and Relative Weighted Product (weighted inpatient workload)	M2	CHAMPUS workload reports with visit weight equal to one
Readiness Measures: Readiness Costs (subtracted from expenditures)	MEPRS G-code	None
GME (Training): Majority of costs of training subtracted from expenditures	MEPRS FAM-code	None

Cost Models

While the DEA models do not require variable transformations, the cost models require them. These statistical models and the variable transformations involved in building them are now discussed.

This dissertation uses variables that are similar to those selected by West (2000) as shown in (2), although there are some notable differences. West's completely specified model (in translog form) is provided below. The variables are categorized into four basic groups: volume, complexity, capacity, and demand uncertainty.

$$C_t = [1 \;\; A_t \;\; V_t \;\; A_t{*}V_t \;\; A^2 \;\; V^2 \;\; Z \;\; S \;\; M \;\; I_t \;\; O_t \;\; A_{t\text{-}1} \;\; V_{t\text{-}1} \;\; R_{t\text{-}1} \;\; Q_{t\text{-}1}][\beta] + \varepsilon_t \qquad (61)$$

where C_t is log(*Cost)* at time t, *A* is log(admissions), *V* is log(visits), *A*V* is the interaction term, *Z* is the log(case mix index), *S* is the log(number of services offered), *M* is a binary indicator variable determining MEDCEN status, *I* is the log(inpatient capacity variable), *O* is the log(outpatient capacity variable), *R* is the log(inpatient coefficient of variation), and *Q* is the log(outpatient coefficient of variation). West's definitions of the inpatient and outpatient capacity variables were loosely based on demand. His volume and complexity variables are straightforward and typical (*A* and *V*; *C, S,* and *M*), while his measures for capacity and uncertainty are contrived yet reasonable. A discussion of these variables is omitted.

Similarly, the transformed variables used for the cross-sectional SFA models in this dissertation are readily expressed as follows.

$$Cost = f\,(Volume,\ Complexity,\ Efficiency) + composed\ error \rightarrow$$

$$\begin{gathered}\text{SFA 1: HN (Half Normal)}\\ C=[1\ \ W\ \ P\ \ Q\ \ A\ \ M\ \ M{*}W][\beta]+v+u\\ \varepsilon=u+v\ \text{(error and random noise)}\\ v\sim N(0,\sigma_v^2)\ \text{(random noise)}\\ u\sim N^{+}(0,\sigma_u^2)\end{gathered}\tag{62}$$

u is inefficiency.

$$\begin{gathered}\text{SFA 2: TN (Truncated Normal)}\\ C=[1\ \ W\ \ P\ \ Q\ \ A\ \ M\ \ M{*}W][\beta]+v+u\\ \varepsilon=u+v\ \text{(error and random noise)}\\ v\sim N(0,\sigma_v^2)\ \text{(random noise)}\\ u\sim N^{+}(\mu,\sigma_u^2)\end{gathered}\tag{63}$$

u is inefficiency.

where C is the log(adjusted cost metric), W is the logarithm of the Bartlett's measure of workload (a linear combination of outpatient and inpatient workload adjusted for complexity), P is the logarithm of a population variable composed of a linear combination of health plan enrollment and population supported variables, Q is the quality variable as in the DEA analysis, A is the access variable as in the DEA analysis, M is an indicator variable for MEDCEN status (1 if MEDCEN, 0 otherwise), $M*W$ is the interaction term for MEDCEN and workload, v is the statistical error term distributed around zero with constant variance, $\varepsilon = u+v$ is the total error (noise and inefficiency, and u is the inefficiency term (time invariant), which is assumed to be either half-normal as in the first model or truncated normal as in the second.

The multi-year panel model varies only by the inclusion of an additional dummy variables *T*. Because of the short panel size (two years) it was impossible to use feasible GLS for the estimation of autoregressive models. The Lee and Schmidt (1993) panel series formulation (56) was therefore employed. The formulation for this specific problem follows.

SFA Panel Data Model

$$C = [1\ W\ \ P\ \ Q\ \ A\ \ T\ \ M\ \ M{*}W][\beta] + v + u \tag{64}$$

$$\varepsilon = u + v \text{ (error and random noise)}$$

$$v \sim N(0, \sigma_v^2) \text{ (random noise)}$$

$$u \sim N^+(0, \sigma_u^2) \text{ and } u \sim N^+(\mu, \sigma_u^2) \text{ (separate models)}$$

A synopsis of the variables and the associated transforms is provided in Table 10.

Table 10. Transformations of Independent Variables.

The table below reflects the transformed independent variables used to predict the dependent variable "budget". The variables are similar to those used by West (2000); however, there is a deliberate attempt to determine efficiency via stochastic modeling. A discussion of the transformations is provided in the forthcoming "Variable Transformations" section.

$W_i = Workload_i = \log(f(RVU_i + RWP_i) + 1.5).$

$P_i = Population_i = \log(f(Population_i + Enrollment_i) + 1.5).$

$Q_i = Prevention_i$, a proxy for quality from DoD Survey.

$A_i = \log(Access_i)$ from Health Care Survey of DoD Beneficiaries.

$E_i = Efficiency$

$M_i = MEDCEN_i \in \{0,1\}$

$Time2 \in \{0,1\}$

$Time3 \in \{0,1\}$

$M_i{*}W_i = MEDCEN_i * W_i$, the interaction

$f \equiv$ factor score from factor analysis

In all models, the volume and complexity drivers are combined into the term W by use of RWP and RVUs. A full discussion of the combinatorial methodology is in Chapter 5; however, this term is an improvement over the traditional count of visits and inpatient workload, as each outpatient visit has varying degrees of complexity that is embedded within the variable. By using the Medicare standard RWP and RVU workload count, workload and complexity are standardized. A more thorough discussion of the SFA model is provided in Chapter 5.

The traditional linear models are solved through OLS and are similar in construction to the SFA models except for the inclusion of the access variables and efficiency scores from the DEA analysis. Specifically, two separate formulations of models provided the basis for analysis. The first model includes the Pareto-Koopmans efficiency indicator, which are determined in a two-stage manger as follows. In stage 1, DEA is applied to determine which DMUs are allocatively efficient. In stage 2, these results are included in the form of a dummy variable in a regression model with *E=1* for efficient observations and *E=0* for inefficient observations as shown in the following model

$$
\begin{gathered}
\text{Linear 1: L1} \\
C = [1\ W\ P\ Q\ A\ T\ M\ M{*}W][\beta] + [E\ \ E{*}W\ E{*}P\ E{*}Q\ E{*}A][\gamma] + \varepsilon \\
E \equiv \text{Pareto Koopmans Efficiency Indicator}, \\
\text{where } E_i = \begin{cases} 0 \text{ if not efficient} \\ 1 \text{ if efficient} \end{cases} \\
M_i = \begin{cases} 0 \text{ if not MEDCEN} \\ 1 \text{ if MEDCEN} \end{cases}
\end{gathered} \tag{65}
$$

where all variables are as in the SFA models and *E* is the vector of Pareto-Koopmans efficiency scores from the DEA analysis (coded 1 if efficient). Only two-way interaction effects are evaluated to retain a parsimonious model.[5]

An additional linear model is also evaluated in this dissertation. The model consists of an Ordinary Least Squares regression as follows:

[5] This model follows the Bardhan, Cooper, and Kumbhakar (1998) model but also includes dummy variables for time.

$$
\begin{gathered}
\text{Linear 2: L2} \\
C = [1\ W\ P\ Q\ A\ E\ T\ M\ M*W][\beta] + \varepsilon \\
E = \text{Technical Efficiency} \\
M = \begin{cases} 0 \text{ if not MEDCEN} \\ 1 \text{ if MEDCEN} \end{cases}
\end{gathered}
\tag{66}
$$

where the vector E is nothing more than the efficiency scores generated from the BCC analysis.

Finally, Bayesian models are proffered and investigated to compare with the SFA and linear models. These models allow for traditional, robust, and ridge regression and are fully developed in the next chapter. The models allow for the predictors to be correlated or even random.

$$
\begin{gathered}
\text{B1: Linear Regression} \\
C = [1\ W\ \ P\ \ Q\ \ A\ \ T\ M\ M*W][\beta] + \varepsilon \\
\beta_i \sim N(0, 10^4)
\end{gathered}
\tag{67}
$$

$$
\begin{gathered}
\text{B2: Ridge Regression} \\
C = [1\ W\ \ P\ \ Q\ \ A\ \ T\ M\ M*W][\beta] + \varepsilon \\
\beta \sim N(0, \eta) \\
\eta \sim IG(10^4, 10^4) \\
\textit{Initializations Vague}
\end{gathered}
\tag{68}
$$

$$\begin{gathered}\text{B3: Robust Regression}\\ C = [1\ W\ \ P\ \ Q\ \ A\ \ T\ M\ M{*}W][\beta] + \varepsilon \qquad (69)\\ \beta \sim MVN(\mu, \Psi)\\ \mu \sim MVN(\gamma, \Lambda)\\ \Psi \sim Wishart(\Omega, 7)\\ \textit{Initializations Vague}\end{gathered}$$

In B1, the β parameters are identically and independently distributed with mean zero and large variance. The variance parameter is updated through sampling. For B2, the coefficients are exchangeable (random variance). In B3, the full random effects model is shown. The regressor terms are considered to be random variables themselves. This formulation is identical to that proposed by Birkes and Dodge (1993). With model descriptions now sufficiently complete, one may now turn to descriptive statistics and variable transformations.

Descriptive Statistics for Military Hospitals

Descriptive statistics for all three years of untransformed data are provided in Table 11. The median facility received \$116.05 million adjusted dollars, and the median population supported was around 58,690 as shown in row 14. The average number of Full Time Equivalent (FTE) employees for each facility was 19,710 as shown in row 8.

Table 11. Descriptive Statistics, Military Hospitals, Untransformed Variables.

The budget variable is measure in millions of dollars. The population, enrollment, RWPs, and RVUs are measured in thousands. The quality / prevention score is based on a percentile, and the access score derives from the HCSDB survey item *Getting Needed Care.*

Variable	N	Average	Std Dev	Minimum	Median	Maximum
1. Access Proxy	72	66.71%	4.08%	57.29%	66.06%	74.79%
2. Efficiency	72	95.53%	3.81%	84.69%	95.76%	100.00%
3. Satisfaction	72	75.05%	4.52%	63.02%	74.08%	85.49%
4. Prevention (Quality Proxy)	72	87.38%	3.71%	78.89%	87.46%	96.38%
5. Cost in millions	72	$ 166.01	$ 130.81	$ 35.05	$116.05	$ 591.70
6. GME$ in millions	72	$ 2.01	$ 3.49	$ -	$ -	$ 12.86
7. Readiness $ in millions	72	$ 5.48	$ 5.34	$ 0.77	$ 4.42	$ 29.46
8. FTEs in 000s	72	19.71	14.15	5.60	14.30	60.34
9. Visits in 000s	72	461.93	254.47	84.75	379.63	908.70
10. RVUs in 000s	72	384.07	244.61	60.06	328.22	957.51
11. RWPs in 000s	72	33.95	36.30	0.96	16.07	127.37
12. Admissions in 000s	72	4.94	4.01	0.10	3.60	13.55
13. Bed Days in 000s	72	17.76	18.43	0.29	9.63	70.84
14. Population in 000s	72	72.85	43.56	13.76	58.69	167.51
15. Enrollment in 000s	72	41.03	25.16	10.43	34.06	105.70

The distribution of the variables in many cases is clearly not normal. A display of histograms illustrates this point (Figure 19).

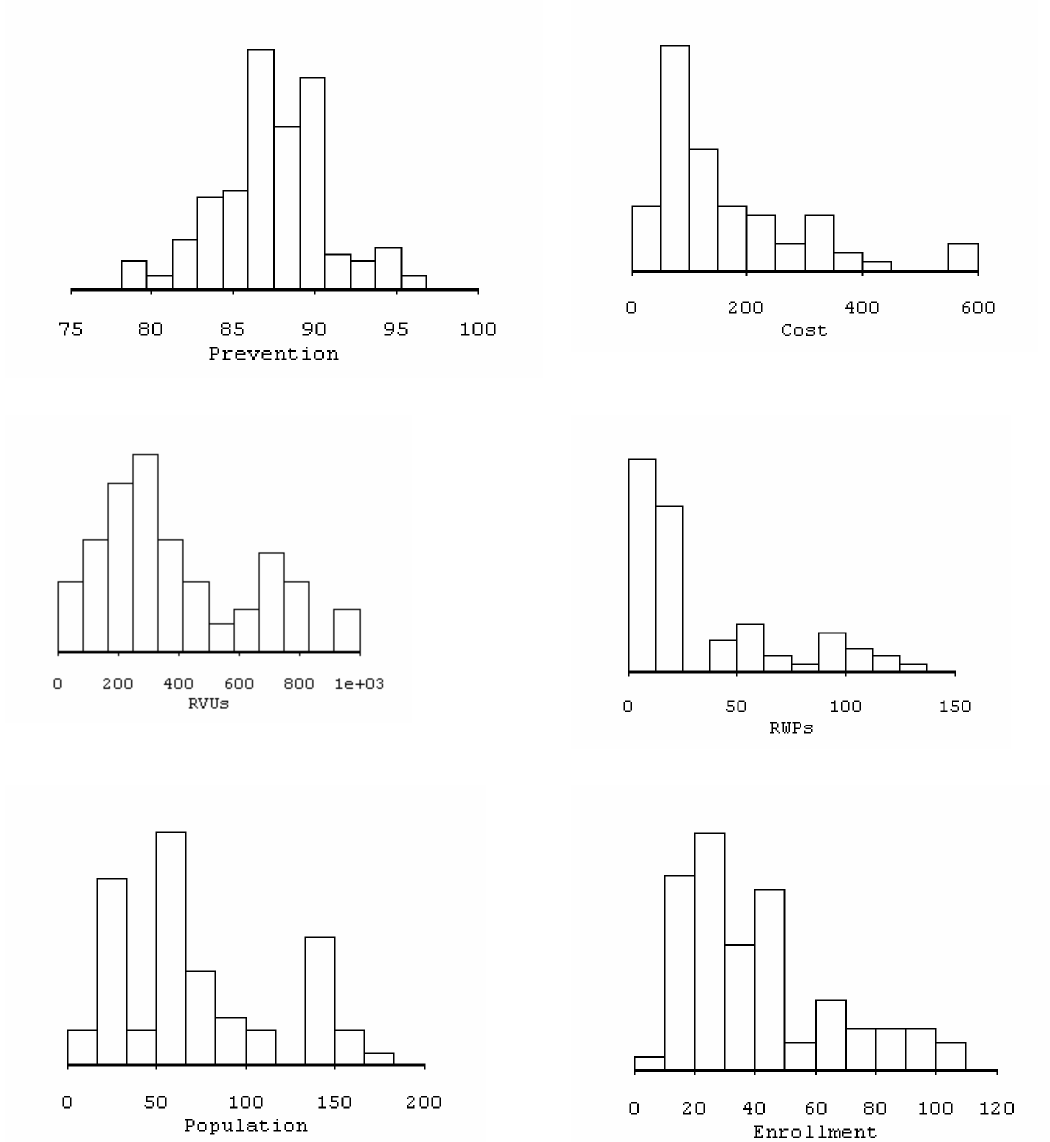


Figure 19. Histograms of Major Variables.

The histograms of some of the major variables of interest suggest lack of normality for cost, RVUs, RWPs, population, and enrollment.

Variable Transformations

In order to accommodate the linear models (including SFA and the resource allocation models), certain variables were combined into factors to reduce collinearity. While DEA requires no parametric assumptions, these assumptions are important for inferential tests of normal linear models.

Investigating the predictor term correlation matrix for all three years of data reveals a high degree of correlation between population and enrollment and also between RVUs and RWPs (see Table 12). The correlation of these predictors is so high that the end result could be highly unstable coefficients, as the $X^T X$ matrix (which is required for the estimation of the parameters) would be near singular.

Table 12. Correlation of Predictor Terms.

The predictor term correlation matrix data reveals statistically significant relationships between several terms (bold); however, the strong relationship between enrollment and population and between inpatient and outpatient workload might be problematic if left untended. The relationship between population and RVUs might also prove problematic.

	1	2	3	4	5	6
1. Population	1.000					
2. Enrollment	**0.858**	1.000				
3. RWPs	0.590	0.319	1.000			
4. RVUs	**0.826**	0.626	**0.878**	1.000		
5. Access	-0.144	-0.294	0.056	-0.034	1.000	
6. Prevention / Quality	0.083	-0.212	0.324	0.280	0.144	1.000

At .858, the correlation between population and enrollment was very likely to be problematic. A simple method for combating the collinearity was to employ factor analysis. Using factor analysis, a single factor (the largest principle component) was generated for each grouping of DMUs that was to be analyzed. This technique led to different results for each DMU grouping. For example, the analysis of Year 2001 data required a different factor analysis then the analysis of Year 2001-2002 data. Separate factor scores and the appropriate coefficients were retained for each analyzed group, so that all parameter estimates required for forecasting in year $t+1$ were obtained in year t.

For the year 2001 data, the factor score accounted for 94% of the variance. Similar results were obtained for all groups. Factor scores were generated using Bartlett's method. The coefficients were saved for forecasting 2002 data, so that the equation for translating population and enrollment into a single factor was available. For year 2001, the equation for determining the population factor was as follows:

$$\textit{Population Factor} = -1.708 + .002 * \textit{Enrollment} + .001 * \textit{Population}. \quad (69)$$

For outpatient and inpatient workload, the correlation was .878. Once again, scores and coefficients for each analytical group were retained. For 2001 data, simple factor analysis produced a single factor also accounting for 94% of the variance. Again, Bartlett's method was employed. For 2001, the equation for determining the workload factor was as follows:

$$\textit{Workload Factor} = -1.297 + .0140 * \textit{RWPs} + .002 * \textit{RVUs}. \quad (70)$$

Location transforms were then performed on both of these factors to make the factor score vectors positive definite in order to investigate a wider array of normalization techniques. {Both the population and workload variables were shifted to the right 1.5 units.} Box-Cox transformations using nonlinear optimization (Newton's method) were investigated for all variables (dependent and predictor) in order to attain a linear mean function and homoskedastic variance. As one might expect from the previous histograms, workload, population, and budget variables were normal under logarithmic transformation. The access predictor term was transformed logarithmically as well. The quality / prevention metric was recoded to eliminate bimodality, as it was resistant to Box-Cox group transformation. The recoded quality / prevention metric follows.

$$Quality = \begin{cases} 70 \geq Q < 75 \rightarrow T(Q) = 1 \\ 75 \geq Q < 80 \rightarrow T(Q) = 2 \\ 80 \geq Q < 85 \rightarrow T(Q) = 3 \\ 85 \geq Q \leq 90 \rightarrow T(Q) = 4 \\ 90 \geq Q \leq 95 \rightarrow T(Q) = 5 \\ 95 \geq Q \leq 100 \rightarrow T(Q) = 6 \end{cases} \qquad (71)$$

So Q, the quality/prevention metric is a step function.

Figure 20 provides the normalized scatterplot using the near optimal λ's. (*λi =0 for all i).* One should note that the Likelihood Ratio Test of the null hypothesis that the distribution is Multi-Variate Normal (MVN) versus the alternate that the distribution is not MVN is significant for the vector of selected λ's ($\lambda_{C,\ P, W, A, Q}$ = [0 0 0 0 0], L = 13.445, p= .02), so the selected transformation is not perfect. The optimal λ's (those which minimize the Likelihood Ratio) are actually $\lambda_{C,\ P, W, A, Q}$ = [-.27 .16 -.15 1.06 1.64], L = .01, p= 1.00; however, the logarithmic transformations are justifiable, a large improvement, and a reasonable simplification.

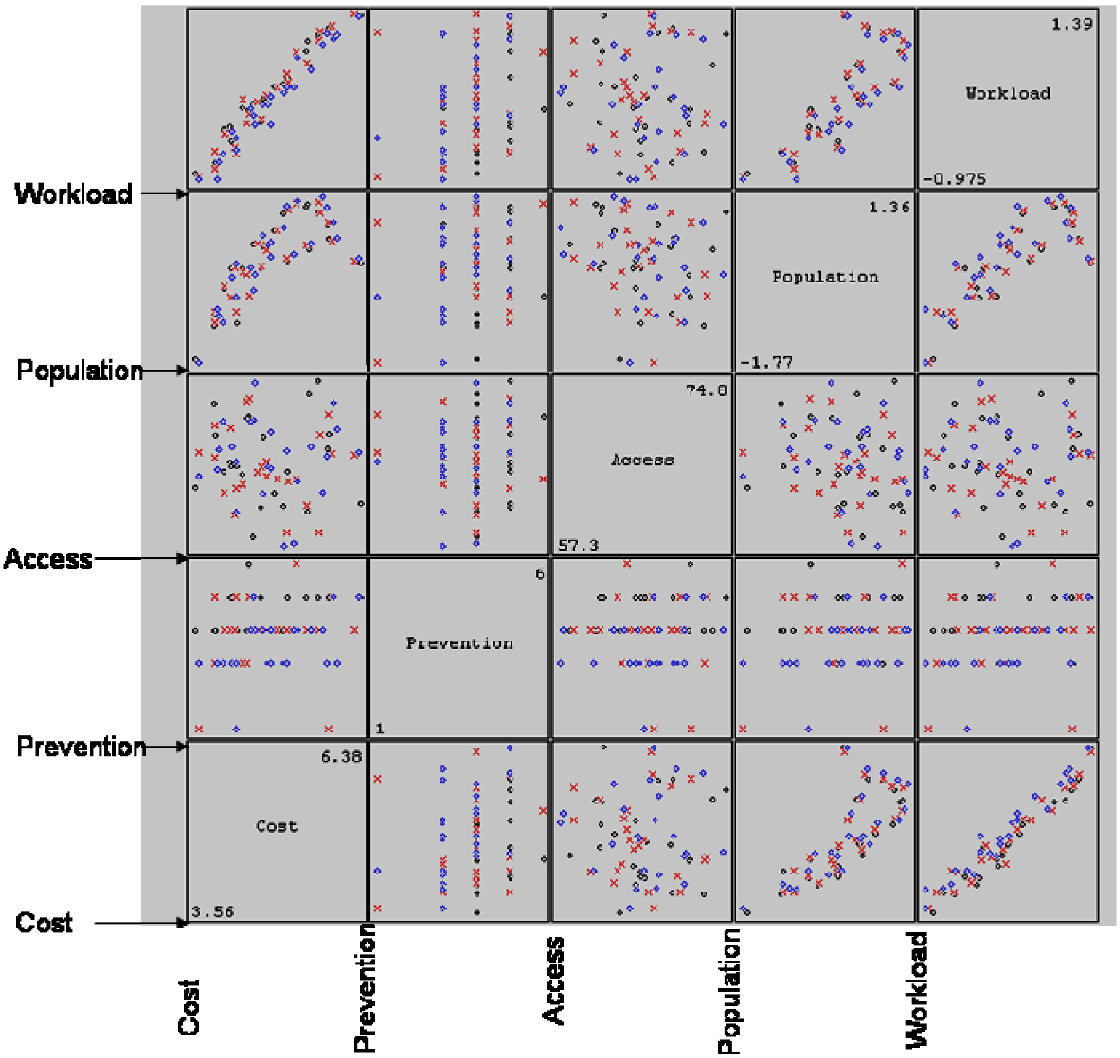


Figure 20. Scatterplot Matrix for Transformed Variables.

The scatterplot matrix after normalization results in a large improvement in the assumption of a multivariate normal mean function. The points are color-coded by year, where the black points represent 2001, the blue points represent 2002, and the red points represent 2003.

Table 13 illustrates the transformation of all variables. These transformations were applied to both the DEA inputs and outputs and used in the cost models. The transformations were logarithmic for all variables with the exception of the quality variable.

Table 13. Variable Transformations.

The variable transformations for all models follow. These transformations ensure normality and prevent issues associated with collinearity (such as instability of coefficient direction and magnitude).

C: *Adjusted budget in 2003 dollars →*	***C:*** *Log of the adjusted budget*
P: *Population,* ***E:*** *Enrollment →*	***P:*** *Log of (factor score + 1.5)*
O: *Outpatient,* ***I:*** *Inpatient Workload→*	***W:*** *Log of (factor score + 1.5)*
Q: *Quality proxy →*	***Q:*** *Recoded quality proxy*
A: *Access→*	***A:*** *Log(Access variable)*

After transformation, the joint distribution of the response variable and regressor terms has a linear mean function, which does not guarantee multivariate normality but is a requirement for the same. All variables were normalized by subtracting their mean and dividing by their standard deviations. This normalization allowed for comparison of effect size in regression by making these values independent of the units of measure employed and was important for model consistency, as in ridge and robust regressions, which rely on standardization. The descriptive statistics for the transformed and standardized variables (all years) are provided in Table 14.

Table 14. Descriptive Statistics for Transformed Variables.

The descriptive statistics for the transformed and standardized variables follow.

	Mean	Median	Stdev	Min	Max	Range
Access	0	0.003	1	-2.297	2.008	4.305
Quality / Prevention	0	0.101	1	-3.003	2.170	5.173
Workload	0	-0.060	1	-1.774	1.811	3.585
Population	0	0.071	1	-2.508	1.577	4.085
Cost	0	-0.131	1	-1.791	2.129	3.920

Summary and Conclusions for Chapter 4

In this chapter, the data sources, models, variable transformations, and descriptive statistics relevant to this research have been detailed. A half-normal and truncated normal SFA model were detailed along with two separate linear regression models (one with Pareto-Koopman's efficiency coded as a binary variable and one with technical efficiency kept as a continuous variable.) Further proposed are three other Bayesian hierarchical models to be solved by Gibbs sampling. In the next chapter, the results of the efficiency analyses are investigated to provide some insight into the Army hospital system.

Chapter 5: Results of Efficiency Analysis, Military Hospitals

DEA Efficiency Analysis, 2001

The BCC-I (Input Oriented) model with non-discretionary inputs as shown in Chapter 3 Equation (7) was used for the year 2001 analysis. One should recall that, in this formulation, the inputs are subdivided into two separate sets. One set accounts for the discretionary inputs, while the other set accounts for the non-discretionary inputs. For consistency between models, variable transformations for the statistical regressions were also used in the DEA analysis.

The input oriented model (7) is translation invariant in the outputs and the non-discretionary inputs. While BCC-O (output oriented) models are not translation invariant in the output axis, the ordering of the DMUs should be highly correlated with the BCC-I models. With this consideration in mind, BCC-O models [see (6b) for reference] were also run to compare efficiency scores and slacks. Since managers may seek either to reduce inputs, while maintaining minimum output levels (i.e., reallocation) or seek to increase outputs using no more than the selected inputs (i.e., process improvement) both orientations were evaluated for comparison. The input and output selection, data sources, and categorization as discretionary or nondiscretionary are provided in Table 9 of Chapter 3. In general, cost and population are the primary inputs, while workload, quality, and access variables are the primary outputs.

The BCC models were implemented using multiple software programs. The model was initially programmed using two different codes in the General Algebraic Modeling System (GAMS) software (www.gams.com). The first code involved establishing two submodels, one for phase 1 analysis (technical efficiency determination) and another for phase 2 analysis (allocative efficiency determination accomplished through the maximizing of the sum of the input excesses and output shortages). The second code used a potentially unstable, single-phase procedure, which assigned to the non-Archimedean element (η in the model provided previously) a value of 10^{-4} (Ali and Seiford, 1993). This procedure was chosen to investigate the DEA model developed by Dyson, Thanassoulis, and Boussofiane (2004) of the Warwick Business School. (See http://www.gams.com/modlib/libhtml/dea.htm for a description and downloadable code). Results were checked independently by use of both the Efficiency Management System (EMS) software provided by Holger Scheel of the University of Dortmund (http://www.wiso.uni-dortmund.de/lsfg/or/ scheel/ems/), and DEA Solver software from Saitech (http://www.saitech-inc.com/Products/Prod-DSP.asp). The use of the DEAP freeware from the University of New England in Australia (http://www.uq.edu.au/economics /cepa/deap.htm) was not possible, as DEAP does not allow for the inclusion of non-discretionary variables. While the technical efficiency scores and slacks output by all the codes were quite similar only DEA Solver and both of the GAMS formulations were identical within rounding error for technical efficiency scores and slacks (see Table 15 for correlations between the results). One should expect only nominal error associated with significant digit changes; however, this is not the case. The implication is that the EMS algorithm is not identical to the traditional DEA model formulation; therefore, DEA Solver and GAMS models were used exclusively in this dissertation. Table 15 provides the Pearson correlations between the technical efficiency scores during a test run of the algorithms, and a further analysis is provided in Appendices 1-4. (Note: the input and output selection for the trial runs is different from that finally selected for this dissertation.)

Table 15. Correlation between DEA Analysis Tools.

The correlation between the four different analyses reveals what appears to be an inconsistency in the technical efficiency scores calculated by EMS and those calculated by DEA Solver and the GAMS formulations.

Table of correlations

	EMS	DEA Solver	1-Phase GAMS	2-Phase GAMS
EMS	1.000			
DEA Solver	0.985	1.000		
1-Phase GAMS	0.985	1.000	1.000	
2-Phase GAMS	0.985	1.000	1.000	1.000

<u>Results of the DEA Analysis, Military Hospitals, Year 2001</u>

Analysis of the year 2001 data proceeded as if there was no knowledge of the data in years 2002 and 2003. The results of the analysis for the year 2001 reveal only small distinctions among the DMUs (see Table 16). This result is expected due to the small number of units included in the analysis and the five inputs and outputs. The number of DMUs should generally be greater than the maximum of the following set: {inputs * outputs, 3(inputs + outputs)}. In this case, a minimum of 15 DMUs is required (24 are available); however, there is little differentiation when the number of DMUs is small (Cooper et al, 2000).

Table 16. Descriptive Statistics for DEA Analysis of 2001.

The descriptive statistics for the DEA technical efficiency scores are provided below.

	Mean	Median	Standard Deviation
Efficiency-Input Orientation	.976	.995	.034
Efficiency-Output Orientation	.989	.997	.014

While the technical efficiency scores provide some idea of the DMU's efficiency, an entity is not Pareto-Koopmans efficient unless it has both a technical efficiency score of one and zero slack (Cooper et al, 2000). For this particular dataset, all DMUs that were evaluated as technically efficient also had the max slack solution equal to zero in both the input and output oriented models (see Table 17). Of the 24 facilities analyzed, 12 of them were determined to be both technically and allocatively efficient for the input and output oriented models. Of these facilities, five are MEDCENs in both analyses. Due to confidentiality, facility names are omitted; however, the names range from the letter A through the letter X. Facilities D, G, N, S, T, V, and X comprise the MEDCENS.

Table 17. Efficient DMUS, Year 2001.

The efficient DMUs for 2001 are shown. MEDCENs are over-represented in this group (five out of seven efficient MEDCENs versus six out of 17 efficient ACHs for the input-oriented analysis). The separate analyses of the MEDCENs and the hospitals will provide additional insight into the validity of grouped analysis.

Input Oriented	Output Oriented
A	A
D (MEDCEN)	D (MEDCEN)
G (MEDCEN)	G (MEDCEN)
J	J
L	L
M	M
N (MEDCEN)	N (MEDCEN)
Q	Q
	R
T (MEDCEN)	T (MEDCEN)
U	U
X (MEDCEN)	X (MEDCEN)

An analysis of those adjudged efficient reveals little similarity. The facilities are spread across a vast geographic area and encompass all sizes and workload levels. The MEDCENs are over-represented, however, which will lead to additional analysis. Those facilities determined Pareto-Koopmans efficient by the input-oriented model were also deemed efficient by the output-oriented model. The consistency of the analysis is reassuring.

Of the 12 inefficient facilities (i.e., DMUs with efficiency scores less than unity or nonzero slacks), only four were determined to be less than 95% efficient in the input oriented model. (None were less than 95% efficient in the output oriented model.) The slacks associated with the inefficiency suggested that cost, quality, and access might be improved for these facilities (see Table 18).

Table 18. DEA Slacks for 2001 Analysis.

The initial DEA analysis for 2001 identified four facilities below 95% technical efficiency. The cost slack is actually the projection, (1-efficiency score)* times the original input, as there is only nondiscretionary input. The units of measure for this table are defined in Table 13.

DMU	Score	Cost Slack	Workload Slack	Quality Slack	Access Slack	Inefficiency ($)
I	0.88	-0.54072	0	0.71	0.06	$ 0.58
K	0.89	-0.47781	0	0.30	-	$ 0.62
P	0.94	-0.25293	0	0.67	-	$ 0.78
F	0.94	-0.28612	0	-	0.02	$ 0.75
H	0.96	-0.20451	0	0.21	-	$ 0.82
V	0.96	-0.18473	0	-	-	$ 0.83
C	0.96	-0.17381	0	0.33	0.01	$ 0.84
O	0.97	-0.14562	0	0.60	-	$ 0.86
E	0.97	-0.13873	0	0.84	-	$ 0.87
W	0.98	-0.11298	0	0.06	0.10	$ 0.89
S	0.98	-0.11336	0	0.21	0.09	$ 0.89
B	0.99	-0.00433	0	0.91	-	$ 0.96

Looking at the bottom four DMUs (I, K, P, and F), one similarity emerges. Three of the four facilities (all but facility I) are located in the northeastern region of the country. There might be a regional factor for low-performing outliers. Further, the bottom four facilities are all ACHs, although one (facility F) is quite large in comparison to the others. There are no other obvious consistencies in the facilities determined to be less than 95% efficient by the input oriented analysis. The cost adjustments recommended by the DEA analysis (in millions) are provided in the last column and are equivalent to the exponential cost slack variable. Based solely on this analysis, there would have been a $10 million reduction for the low-performing facilities for the Year 2001 without affecting performance. Considering the enormity of the budget of these facilities ($4 billion), these slacks seem to be reasonable in terms of magnitude. Facilities E and B both have large quality slacks, indicating that these

DMUs could have performed nearly one point better on the adjusted quality scale based on the resources provided. Access slacks are nominal throughout.

Another analysis of the slacks was then performed using a RAM model (see Appendix 5 for the GAMS replication code). The RAM model is non-oriented and may provide interesting results for resource allocation models. One would expect the results to support some of the findings in the BCC-I model. Table 19 provides the results.

Table 19. RAM Results, Year 2001.

The RAM formulation identifies a subset of the BCC inefficient DMUs; however, the only cost inefficient facilities are determined to be K and S. These values may be more consistent with Table 19 than apparent at first glance, because of the inferior representation of θ^* in Table 19 with reference to the slack values. Specifically, 1-θ^* is distributed evenly across the slacks without dictating the mixture.

DMU	C Slack	W Slack	Q Slack	A Slack	Translated $
C			0.972		
E		0.106	0.978		
H			0.958		
I		0.397	1.161	0.039	
K	0.336		0.701	0.022	$ 1.40
S	0.107	0.002	0.87		$ 1.11
V		0.075	0.207		

All facilities identified by RAM as having slack resources were also identified as inefficient in the BCC analyses. About $2.51 million of excess cost was found by RAM, whereas the BCC-I model discovered $9.69 million. The RAM model suggests that Facility I has the largest workload slack, while Facility H has the largest quality slack. Access slacks are nominal in both the RAM and BCC results. From both the BCC and the RAM analysis, Facilities I and K are clear outliers.

SFA Efficiency Estimates

In order to avoid "methodological bias," (Charnes, Cooper, and Sueyoshi ,1988) it is important to explore techniques other than DEA for evaluation of efficiency. Stochastic Frontier Analysis was an appropriate secondary tool widely used in the healthcare setting.

As discussed previously, Stochastic Frontier Analysis (SFA) is a commonly used method for evaluating efficiency. Unlike DEA, SFA requires a priori model selection and is highly sensitive to collinearity. Therefore, a comparison based on the set of variables included in the DEA model is not feasible, as there is significant and strong collinearity among many of the variables. A reduced variable set is used. The typical time invariant SFA Model is provided in (47). Recall that the general formulation of the model is $y_{it} = f(X_{it},\beta)exp(v_{it}+u_i)$ for a cost frontier.

While DEA directly evaluates efficiency in terms of the use of outputs and inputs, SFA evaluates inefficiency through parameter estimation of a cost frontier model. For the SFA analysis, the analysis of a cost frontier was performed using the following general equation: *cost = f(outputs)*, where cost is defined as the actual expenditures of the facilities as (13) in Chapter 3. The cross-sectional models investigated for the year 2001 are represented in (62) and (63) of Chapter 4.

Comparison of DEA, SFA, and COLS Estimates

Efficiency analyses were conducted using the LIMDEP software (Greene, 2002; www.limdep.com). LIMDEP is an econometric software program that provides the ability to evaluate both cost and production frontiers with multiple error formulations. Maximum Likelihood Estimates (MLE) were obtained for all parameters in both the half-normal and truncated normal formulations. Only the workload term is statistically significant in both analyses. Corrected Ordinary Least Squares efficiency scores were also generated. Tables 20 and 21 summarize the results of the DEA, SFA, and COLS analyses.

Table 20. Descriptive Statistics: DEA, SFA, and COLS, 2001.

Descriptive statistics for the SFA models and the DEA models are shown side-by-side. It is clear to see that there is little difference between the half normal and truncated normal formulations but that these formulations are not of the same magnitude as the DEA model.

	Mean	Median	Std Dev
DEA-I [6]	.975	.987	.034
DEA-O	.989	.997	.014
SFA-HN	.787	.813	.180
SFA-TN	.787	.813	.180
COLS	.756	.730	.118

The technical efficiency results from LIMDEP (one minus the parameter estimates for the inefficiency) are provided side-by-side with the average results produced from the previous DEA analyses and the Corrected Ordinary Least Squares results. The disparities are evident as illustrated in Table 22.

[6] DEA-I is the BCC-I (input oriented) results, while DEA-O is the BCC-O (output oriented) results. SFA-HN is the Half-Normal SFA model. SFA-TN is the Truncated Normal SFA model. COLS is the Corrected Ordinary Least Squares model.

Table 21. Efficiency Score Comparisons: DEA, SFA, and COLS, 2001.

The side-by-side comparison of the efficiency scores reveals that the top and bottom performers are largely the same for all analyses. Both the input-oriented DEA scores (DEA-I) and the output-oriented DEA scores (DEA-O) are similar in rank ordering. The SFA formulations are also similar when compared together but show some disparity when compared to the DEA results. The COLS scores are more consistent with the SFA-HN and SFA-TN formulations.

DMU	DEA-I	DEA-O	SFA-HN	SFA-TN	COLS	MEAN EFF.
J	1.000	1.000	1.000	1.000	1.000	1.000
X	1.000	1.000	1.000	1.000	0.950	0.990
R	1.000	1.000	1.000	1.000	0.930	0.986
B	0.990	0.995	0.974	1.000	0.880	0.968
N	1.000	1.000	1.000	1.000	0.800	0.960
U	1.000	1.000	0.920	0.990	0.840	0.950
D	1.000	1.000	0.938	0.999	0.800	0.947
L	1.000	1.000	0.917	0.955	0.850	0.944
A	1.000	1.000	0.912	0.940	0.800	0.930
V	0.964	0.984	0.909	1.000	0.770	0.925
S	0.981	0.987	0.867	0.904	0.720	0.892
W	0.975	0.968	0.818	0.788	0.830	0.876
O	0.969	0.985	0.758	0.722	0.720	0.831
G	1.000	1.000	0.737	0.797	0.620	0.831
C	0.965	0.970	0.734	0.697	0.740	0.821
H	0.957	0.978	0.738	0.713	0.710	0.819
R	1.000	1.000	0.676	0.650	0.700	0.805
M	1.000	1.000	0.661	0.650	0.710	0.804
E	0.975	0.972	0.686	0.658	0.700	0.798
T	1.000	1.000	0.599	0.727	0.620	0.789
P	0.940	0.978	0.663	0.606	0.700	0.777
F	0.944	0.971	0.551	0.484	0.680	0.726
K	0.888	0.978	0.458	0.485	0.530	0.668
I	0.885	0.960	0.432	0.413	0.550	0.648

It is quite interesting to see that facilities in the northeastern region of the country are still represented in the bottom portion of the analysis. Specifically, F, K, and P have three of the four lowest mean efficiency scores from all methods and are located in this region. Facility I, which is in the southern United States, lies at the low end of all the methods used. The lowest performing MEDCEN (medical center) in terms of mean efficiency (the average of all the efficiency scores) is Facility T. While the SFA and COLS analyses determined that the facility was inefficient, the DEA analysis suggested that the facility is Pareto-Koopmans efficient.

For the high-performance DMUs, facilities J and X are diverse in that one is a MEDCEN in the Southeast and the other is a MEDDAC in the central region of the country. There is little location or size similarity in any of the top-performing facilities.

The correlation matrix provided in Table 22 reveals positive correlation and statistical significance of the correlation between all models. The SFA half and truncated normal models are nearly identical; however, just as in the Reinhard et al study (1997), there is disparity between the DEA and SFA models with DEA-I exhibiting higher correlation with the other models in comparison with DEA-O . The largest correlation value gap is between the output oriented DEA model and the COLS model.

Table 22. Model Correlation, 2001.

The correlation matrix for the technical efficiency scores is provided below. While highly correlated, there is still significant difference in the estimates between the DEA and SFA models.

	DEA-I	**DEA-O**	**SFA-HN**	**SFA-TN**	**COLS**
DEA-I	1.000				
DEA-O	0.806**	1.000			
SFA-HN	0.729**	0.608**	1.000		
SFA-TN	0.745**	0.683**	0.970**	1.000	
COLS	0.641**	0.482*	0.897**	0.815**	1.000

** *p<.01, *p<.05*

A graph of the Z-standardized efficiency scores, as in Figure 21, exhibits the congruence between the efficiency evaluations with COLS showing the widest variation at the high end and DEA-I accounting for the widest variation at the low end of the efficiency scores.

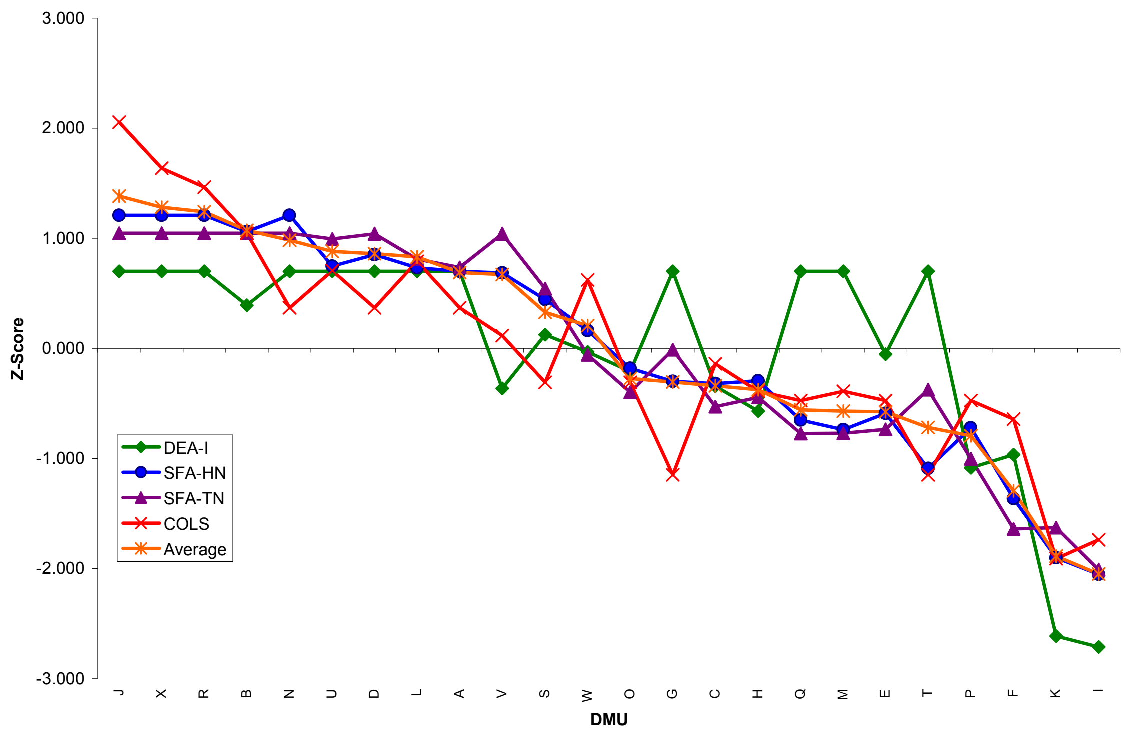


Figure 21. Efficiency Score Comparison, 2001.

The graph of the efficiency scores for four of the five models is provided above. The values are sorted by the average efficiency score denoted "Average." The output oriented model is omitted as it provides little differentiation with the current number of DMUs. Strikingly, the results are similar at both ends of the analysis, with deviations largely near the average.

In general, all efficiency formulations identify some of the same outliers. While the SFA analysis of the cost frontier is interesting, it is dependent upon the assumptions of the underlying model. The DEA analysis requires no assumptions and is certainly more flexible but is sensitive to outliers. One should also note that the COLS model used is a CRS formulation, which may be problematic.

Efficiency Analysis, 2001-2002

It is interesting to repeat this analysis of efficiency for the years 2001-2002 to determine if panel data improves the correlation between methods and if the results identify generally the same efficient and inefficient facilities. The formulation of the BCC models for the year 2001 and 2002 is nearly identical to that for the cross-sectional analysis; however, the data was grouped such that Hospital A in 2001 was considered to be a separate DMU from Hospital A in 2002. This formulation increased the number of DMUs for evaluation from 24 to 48.

Table 23 lists the DMUs that were evaluated as both allocatively and technically efficient. The Year 2001 DMUs are somewhat over-represented, which is suggestive of a decrease in efficiency over time. Both the input and output oriented models suggested the same DMUs as allocatively efficient for both years of the analysis, which was not to be expected.

Table 23. Efficient DMUs for Years 2001-2002.

Facilities L, T, and U were measured as Pareto-Koopmans efficient for both years.

Panel-Series-Both Orientations	
Year 01	**Year 02**
A	L
D (MEDCEN)	M
G (MEDCEN)	N (MEDCEN)
J	R
L	S
N (MEDCEN)	T (MEDCEN)
T (MEDCEN)	U
U	
X (MEDCEN)	

This two-year analysis results in the identification of the same efficient facilities for the Year 2001 as in the previous cross-sectional study (compare Table 17 with Table 23) with the exception of facilities M, Q, and T. As more DMUs are included in the analysis, differentiation becomes better. Only facilities L, N, T, and U are efficient for both time periods. The two MEDCENS are large GME producing facilities in geographically different areas of the country, while Facility U is a small ACH on a training installation, which has a high volume of troop rotation through the installation.

The DMUs evaluated as inefficient are listed in Table 24. This table show the results for each of the two years in the combined analysis, sorted by year and by efficiency score. Facilities C, F, P, and I are present for both years. The cost slack indicates that a total of nearly $26 million might have been reallocated during the two-year period. This amount reflects less than half of a percent of the total $8 billion expended. Some facilities (such as C, X, and B) exhibit significant quality slack. Based upon the resources received, one should expect that these facilities should have scored one point or more higher on the recoded quality / prevention scale.

Table 24. Inefficient DMUs, 2001-2002.

The following facilities were inefficient for the input oriented DEA analysis. Facilities P, K, and I were identified for both years. These facilities are MEDDACS. Bolded DMUs appear as inefficient for both years.

Year	DMU	Score	C Slack	W Slack	Q Slack	A Slack	Trans. $
2001	**I**	0.881	(0.560)	-	0.786	0.056	$ 0.57
2001	**K**	0.887	(0.483)	-	0.290	-	$ 0.62
2001	**P**	0.940	(0.252)	-	0.668	-	$ 0.78
2001	**F**	0.943	(0.289)	-	-	0.024	$ 0.75
2001	**H**	0.946	(0.259)	-	-	-	$ 0.77
2001	**C**	0.948	(0.255)	-	0.238	0.008	$ 0.77
2001	**W**	0.955	(0.203)	-	0.119	0.092	$ 0.82
2001	**V**	0.963	(0.189)	-	-	-	$ 0.83
2001	**O**	0.964	(0.171)	-	-	-	$ 0.84
2001	**Q**	0.968	(0.144)	-	-	-	$ 0.87
2001	**E**	0.970	(0.164)	-	0.816	-	$ 0.85
2001	S	0.978	(0.127)	-	0.077	0.011	$ 0.88
2001	R	0.985	(0.064)	-	0.165	-	$ 0.94
2001	**B**	0.987	(0.054)	-	0.628	-	$ 0.95
2001	M	0.989	(0.043)	0.015	-	-	$ 0.96
2002	**K**	0.884	(0.496)	-	0.429	0.013	$ 0.61
2002	**I**	0.900	(0.472)	-	-	-	$ 0.62
2002	**P**	0.904	(0.409)	-	0.587	0.068	$ 0.66
2002	G	0.923	(0.424)	-	0.607	0.010	$ 0.65
2002	**O**	0.929	(0.339)	-	-	-	$ 0.71
2002	**C**	0.931	(0.346)	-	1.102	-	$ 0.71
2002	**F**	0.934	(0.339)	-	-	-	$ 0.71
2002	**H**	0.942	(0.278)	-	-	-	$ 0.76
2002	**W**	0.947	(0.245)	-	-	-	$ 0.78
2002	**V**	0.948	(0.266)	-	0.440	0.068	$ 0.77
2002	**Q**	0.962	(0.168)	-	-	-	$ 0.85
2002	X	0.963	(0.210)	-	1.000	0.040	$ 0.81
2002	**B**	0.976	(0.099)	-	1.000	-	$ 0.91
2002	J	0.980	(0.080)	-	0.889	0.037	$ 0.92
2002	A	0.988	(0.046)	-	0.097	-	$ 0.95
2002	**E**	0.992	(0.044)	-	0.572	-	$ 0.96
2002	D	0.996	(0.025)	-	-	0.026	$ 0.98

The results of the two-year BCC analysis are generally consistent with the single year analysis; however, one might review whether or not a regional effect is actually present. In the two year study, the inefficient facilities represent more of a cross-section of the hospitals in the AMEDD, so the regional effect disappears.

The DEA analysis indicates that adjustments might be made to improve efficiency by reallocation of funds. The results of the RAM model for both years provides additional insight.

Table 25. RAM Results, 2001-2002.

The results of the RAM analysis indicate that facilities I and K were cost inefficient. Facility K was cost inefficient both years, while Facility I was deemed cost inefficient only in 2002. Facility I was evaluated as workload inefficient for both years. Bolded facilities exhibited quality inefficiency both years.

Year	DMU	C Slack	W Slack	Q Slack	A Slack	Trans. $
2001	**K**	0.261		1.214		$ 1.30
2001	E		0.094	1.019		
2001	**I**		0.186	1.639	0.042	
2001	**V**			0.388		
2001	R			0.587	0.000	
2001	**B**			0.792	0.011	
2001	**H**			0.993		
2001	**C**			1.014		
2001	S			1.113		
2002	**I**	0.019	0.132	0.551		$ 1.02
2002	**K**	0.148		1.794	0.038	$ 1.16
2002	J		0.078	1.016	0.039	
2002	G		0.101	1.526	0.014	
2002	P		0.250	0.437	0.078	
2002	W			0.147		
2002	D			0.177	0.003	
2002	A			0.407		
2002	**H**			1.100		
2002	O			1.179		
2002	**V**			1.393	0.077	
2002	**B**			1.796		
2002	**C**			2.051		

In the RAM formulation, K and I remain inefficient; however, P is not identified as having any slack in 2002. Facilities B, H, V, and I were identified as inefficient in 2001 and 2002 for both the BCC and RAM analysis. The congruence is reassuring.

An analysis of the referent sets for the BCC model is interesting in that it identifies the most commonly occurring peers. The most commonly occurring peer (R02) is actually a MEDDAC in the central region of the country.

Table 26. Peer Groups, 2001-2002 Analysis.

DMU/YR	Frequency
A01	14
D01	4
G01	0
J01	15
L01	1
N01	10
T01	0
U01	7
X01	11
L02	1
M02	2
N02	1
R02	17
S02	0
T02	1
U02	0

Comparison of DEA, SFA, and COLS for Years 2001-2002

With the DEA analysis for the panel data now in hand, SFA models are modified to include time invariant efficiency by introducing a time subscript *t* on the models. Specifically, the SFA model is formulated as before with the addition of dummy variables for time (see Lee and Schmidt, 1993) and estimated by Maximum Likelihood Estimation. The results of the panel analysis provide an estimate of efficiency for each DMU for each time period and are compared to the average efficiency calculated by DEA analyses and the COLS analysis. Based on the apparent similarity of the separate model formulations discovered during the cross-sectional analysis, only the half-normal model is investigated.

The time varying estimates for technical efficiency from the SFA analysis of panel data are displayed side by side with the estimates from DEA and from COLS in the following tables.

As can be seen in Table 27, the SFA and DEA efficiency scores are much closer in terms of magnitude than in the cross-sectional analysis (Table 21) with the COLS estimates further away.[7] This result may be due to the fact that COLS does not use all of the information in the estimates of the

[7] While the magnitude is similar, the rank ordering is not, which will be illustrated in Table 29.

coefficients, as it relies only on the largest outlier and the regression slope. In general, the efficiency scores are quite high.

Table 27. Inefficient DMUS from Panel Series Analysis, 2001.
The analysis is sorted by average efficiency score for clarity.

DMUS	DEA-2	DEA-O	SFA-HN	COLS	Average Efficiency
J 01	1.000	1.000	0.972	0.979	0.988
X 01	1.000	1.000	0.972	0.958	0.982
R 01	0.985	0.993	0.971	0.882	0.958
U 01	1.000	1.000	0.970	0.826	0.949
B 01	0.987	0.992	0.971	0.845	0.949
L 01	1.000	1.000	0.968	0.792	0.940
N 01	1.000	1.000	0.969	0.770	0.935
A 01	1.000	1.000	0.969	0.768	0.934
D 01	1.000	1.000	0.968	0.765	0.933
W 01	0.955	0.964	0.970	0.815	0.926
M 01	0.989	0.997	0.966	0.712	0.916
V 01	0.963	0.983	0.966	0.723	0.909
S 01	0.978	0.985	0.967	0.692	0.905
Q 01	0.968	0.994	0.965	0.688	0.904
O 01	0.964	0.985	0.967	0.691	0.902
C 01	0.948	0.969	0.967	0.705	0.897
P 01	0.940	0.978	0.967	0.701	0.896
H 01	0.946	0.978	0.966	0.677	0.892
T 01	1.000	1.000	0.961	0.590	0.888
F 01	0.943	0.970	0.964	0.669	0.887
E 01	0.970	0.968	0.965	0.644	0.887
G 01	1.000	1.000	0.962	0.585	0.887
I 01	0.881	0.960	0.961	0.542	0.836
K 01	0.887	0.978	0.960	0.519	0.836

Note from the above table that the facilities adjudged least efficient are G, I, and K, where G is a MEDCEN and I and K are ACHs. These are the same facilities identified in the 2002 analysis.

Table 28 apparently confirms what was exhibited in Table 27 with COLS consistently below the efficiency average.

Table 28. Inefficient DMUS from Panel Series Analysis, 2002.

DMUS	DEA-2	DEA-O	SFA-HN	COLS	Average Efficiency
R 02	1.000	1.000	0.973	1.000	0.993
J 02	0.980	0.979	0.971	0.893	0.956
N 02	1.000	1.000	0.970	0.823	0.948
A 02	0.988	0.993	0.969	0.796	0.937
L 02	1.000	1.000	0.967	0.773	0.935
D 02	0.996	0.997	0.968	0.769	0.932
M 02	1.000	1.000	0.965	0.757	0.930
E 02	0.992	0.991	0.968	0.758	0.927
S 02	1.000	1.000	0.967	0.741	0.927
X 02	0.963	0.962	0.969	0.770	0.916
U 02	1.000	1.000	0.968	0.696	0.916
W 02	0.947	0.978	0.968	0.769	0.916
B 02	0.976	0.992	0.969	0.720	0.914
V 02	0.948	0.953	0.968	0.763	0.908
H 02	0.942	0.975	0.967	0.712	0.899
Q 02	0.962	0.998	0.966	0.670	0.899
T 02	1.000	1.000	0.963	0.614	0.894
O 02	0.929	0.971	0.967	0.706	0.893
F 02	0.934	0.974	0.963	0.650	0.880
C 02	0.931	0.970	0.966	0.627	0.873
P 02	0.904	0.957	0.965	0.651	0.869
G 02	0.923	0.957	0.965	0.631	0.869
I 02	0.900	0.978	0.962	0.616	0.864
K 02	0.884	0.968	0.962	0.562	0.844

The correlations Table 29 move in the opposite direction with COLS and SFA represented by a high correlation (.96), while both COLS and SFA exhibit low correlations with DEA. The correlations of the parametric approaches with DEA are lower than that found in Table 23, so the inclusion of time periods and the introduction of time dependent features resulted in the deterioration.

Table 29. Correlation Coefficients, 2001-2002.

The correlation coefficients between the efficiency estimates suggest that there exists some disparity. The most noticeable differences are between the output oriented DEA and the SFA and COLS models. There is a bifurcation between the DEA methods and the residual methods visible in this table.

	DEA-I	DEA-O	SFA-HN	COLS	Average EFf.
DEA-I	1.000				
DEA-O	0.811**	1.000			
SFA-HN	0.576**	0.307*	1.000		
COLS	0.606**	0.391*	0.941**	1.000	
Average Eff.	0.801**	0.607**	0.901**	0.960**	1.000

** *p<.01, *p<.05*

The graph of the efficiency scores shows less variability as might be expected but COLS continues to show the largest variation at the upper end (left side) of the chart, while DEA continues to show the largest variation at the lower end (right side) of the chart as was the case in Figure 21. The graphical comparison of the efficiency scores is based on normalized efficiency estimates (Z-scores).

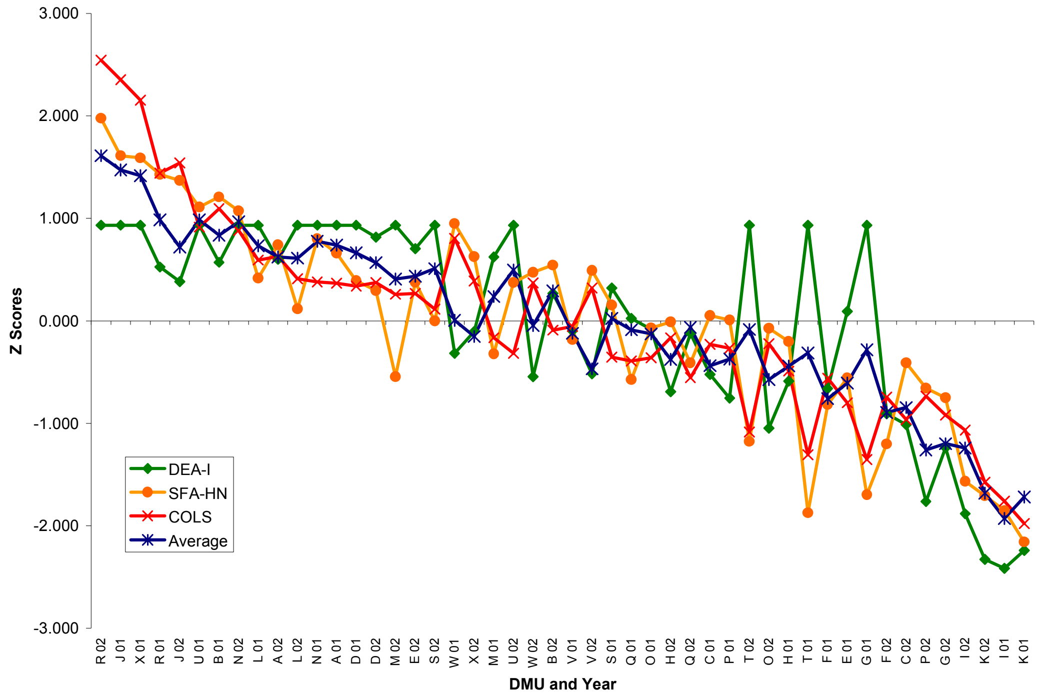


Figure 22. Efficiency Estimates, 2001-2002.
The figure above depicts the estimates of efficiency by each estimator. Near the extremes, the estimators appear to be fairly congruent.

The previous graph reveals an interesting phenomenon when observed closely. While the analyses of efficiency are tight at both extremes as before, the disparity from point T02 to point F02 is more pronounced. Specifically, the SFA and COLS models are opposite of the DEA models at the point where pronounced spikes occur at T02, T01, and G01. These facilities are MEDCENS, which are likely (and in fact do) have high costs. Evaluating the facilities closer, the quality metrics for facility T for both years are the highest and the access measure for the G01 is the highest. The cost formulations attribute less weight to quality and access than DEA.

Efficiency Analysis, 2001-2003

The efficiency analysis for the years 2001-2003 resulted in findings similar to the findings for 2001-2002. Table 30 provides the efficiency scores generated by all analyses sorted by year and by average efficiency score. Further, the table is segmented into thirds to see if the efficiency analyses successfully identify consistent low and high performing facilities. From Table 30, it is apparent that Facilities J, N, R, and D are in the top-third for average efficiency for all three years, which is a finding identical to the two year panel analysis (Table 22). Facilities P, M, I, G, and F are in the bottom third, which is consistent with Table 22 with the exception of Facility G. Facility G was in the middle third for the two-year analysis. Again, the addition of more DMUs allows for more discrimination in the results.

Table 30. Efficiency Analysis, 2001-2003.

The following table shows the efficiency scores sorted by year and by average efficiency. The chart is segmented into thirds and the bolded facilities are those that remained in the tier for all three years.

Year	DMU	DEA-I	DEA-O	SFA	COLS	Avg.	Year	DMU	DEA-I	DEA-O	SFA	COLS	Avg.	Year	DMU	DEA-I	DEA-O	SFA	COLS	Avg.
2001	**J**	1.00	1.00	0.96	0.91	0.97	2002	**R**	1.00	1.00	0.96	1.00	0.99	2003	L	0.96	0.99	0.95	0.92	0.95
2001	B	1.00	1.00	0.96	0.90	0.96	2002	**N**	1.00	1.00	0.96	0.88	0.96	2003	**J**	0.97	0.98	0.96	0.89	0.95
2001	**N**	1.00	1.00	0.95	0.90	0.96	2002	**J**	0.98	0.98	0.96	0.90	0.95	2003	**R**	0.99	0.99	0.96	0.86	0.95
2001	**R**	0.99	0.99	0.96	0.92	0.96	2002	**D**	1.00	1.00	0.94	0.87	0.95	2003	**N**	0.97	0.98	0.95	0.86	0.94
2001	**D**	1.00	1.00	0.95	0.89	0.96	2002	B	0.98	0.99	0.95	0.86	0.94	2003	A	0.95	0.99	0.94	0.87	0.94
2001	L	1.00	1.00	0.95	0.87	0.95	2002	A	1.00	1.00	0.95	0.82	0.94	2003	S	0.94	0.99	0.94	0.88	0.94
2001	X	1.00	1.00	0.96	0.84	0.95	2002	W	0.95	0.98	0.95	0.85	0.93	2003	**D**	0.94	0.99	0.93	0.87	0.93
2001	U	1.00	1.00	0.95	0.84	0.95	2002	H	0.95	0.98	0.94	0.85	0.93	2003	W	0.93	0.96	0.95	0.87	0.93
2001	A	1.00	1.00	0.95	0.81	0.94	2002	L	1.00	1.00	0.94	0.77	0.93	2003	K	0.94	0.98	0.94	0.85	0.93
2001	S	0.98	0.98	0.94	0.81	0.93	2002	**O**	0.94	0.97	0.94	0.85	0.92	2003	U	0.98	0.99	0.95	0.77	0.93
2001	W	0.97	0.96	0.95	0.81	0.92	2002	X	0.97	0.96	0.95	0.79	0.92	2003	B	0.92	0.98	0.94	0.86	0.93
2001	H	0.95	0.98	0.93	0.83	0.92	2002	V	0.95	0.95	0.95	0.82	0.92	2003	H	0.91	0.97	0.93	0.88	0.92
2001	V	0.95	0.97	0.94	0.80	0.91	2002	S	0.97	0.99	0.93	0.77	0.92	2003	X	0.94	0.96	0.95	0.81	0.91
2001	C	0.95	0.97	0.94	0.79	0.91	2002	E	1.00	1.00	0.95	0.72	0.92	2003	**O**	0.95	0.98	0.94	0.79	0.91
2001	**O**	0.96	0.98	0.93	0.77	0.91	2002	C	0.94	0.97	0.93	0.80	0.91	2003	E	0.97	0.97	0.95	0.73	0.91
2001	Q	0.95	0.99	0.92	0.76	0.91	2002	T	1.00	1.00	0.91	0.72	0.91	2003	T	1.00	1.00	0.92	0.69	0.90
2001	**P**	0.96	0.98	0.94	0.74	0.91	2002	**I**	0.89	0.97	0.91	0.85	0.91	2003	**G**	0.90	0.98	0.92	0.80	0.90
2001	**M**	0.96	0.99	0.93	0.73	0.90	2002	**G**	0.93	0.96	0.93	0.78	0.90	2003	**P**	0.89	0.96	0.94	0.79	0.89
2001	K	0.89	0.97	0.89	0.85	0.90	2002	Q	0.96	1.00	0.93	0.71	0.90	2003	C	0.88	0.94	0.92	0.80	0.89
2001	**I**	0.90	0.96	0.90	0.78	0.88	2002	U	1.00	1.00	0.93	0.65	0.90	2003	**I**	0.85	0.97	0.89	0.84	0.89
2001	**G**	0.94	1.00	0.90	0.69	0.88	2002	**M**	0.94	0.98	0.93	0.73	0.89	2003	**F**	0.91	0.98	0.92	0.73	0.89
2001	T	1.00	1.00	0.89	0.64	0.88	2002	**P**	0.90	0.96	0.93	0.78	0.89	2003	V	0.89	0.94	0.93	0.74	0.87
2001	E	0.98	0.97	0.93	0.66	0.88	2002	K	0.92	0.97	0.91	0.74	0.88	2003	**M**	0.91	0.98	0.92	0.68	0.87
2001	**F**	0.93	0.96	0.92	0.73	0.88	2002	**F**	0.92	0.97	0.91	0.73	0.88	2003	Q	1.00	1.00	0.93	0.52	0.86

For the 2001-2003 efficiency analysis, the COLS efficiency estimator was grossly inconsistent with the DEA analysis. The reason may be related to the fact that COLS is a Constant Returns to Scale model, whereas the BCC formulation of the DEA models ensures Variable Returns to Scale. A single outlier may greatly affect the COLS CRS frontier. Previous analysis revealed the presence of an extreme X, Y outlier (Facility T), which serves to shift the COLS frontier significantly. The correlation matrix for all estimators is provided in Table 31.

Table 31. Correlation Coefficients, 2001-2003.

The efficiency estimators for all three years are problematic. Specifically, the COLS estimator is only partially consistent with the SFA estimator and is nearly uncorrelated with the DEA formulations. While the DEA formulations are consistent, the output-oriented model exhibits little linear relationship to the SFA model.

	DEA-I	DEA-O	SFA	COLS
DEA-I	1.000			
DEA-O	0.774**	1.000		
SFA	0.581**	0.269	1.000	
COLS	0.031	0.072	0.513**	1.000

**$p<.01$

It is quite interesting to note that the input or output orientation greatly affects the size effect of the relationship between the SFA and DEA models and that the correlation weakens as the number of DMUs increases. This weakening of the correlation as DMUs increase is logical, as these methods are sensitive and respond differently to outliers. Since both DEA and COLS are more sensitive to outliers than SFA, it is possible to obtain very poorly correlated efficiency analyses. Because there is reason to believe that hospitals exhibit VRS, COLS efficiency scores are not included in any other models.

A closer analysis of the DEA results is appropriate, as these results will be used in linear models with DEA efficiency scores and compared to the Stochastic Frontier cost estimates. First, it is interesting to evaluate each year individually (number of DMUs = n=24 x 3) and compare this analysis to all hospitals measured in all years as a grouped entity (n=72). The results of both analyses are nearly identical as depicted in the next graphs and as noted in the correlation coefficient of .964. In Figure 23, the average efficiency scores for the grouped DEA analysis suggest that Facilities T, U, N, R, and L are high performers, while Facilities G, F, P, K, and I are low performers. The ungrouped analysis of the hospitals (Figure 24) provides somewhat less differentiation due to the loss of the degrees of freedom; however, Facilities G, F, P, K, and I are in the bottom seven facilities based on average efficiency score, while Facilities T, U, N, and R are in the top seven. From all of the analyses performed so far, clear evidence emerges that Facilities P, K, I, and F are inefficient. This result is independent of method used to evaluate the facilities.

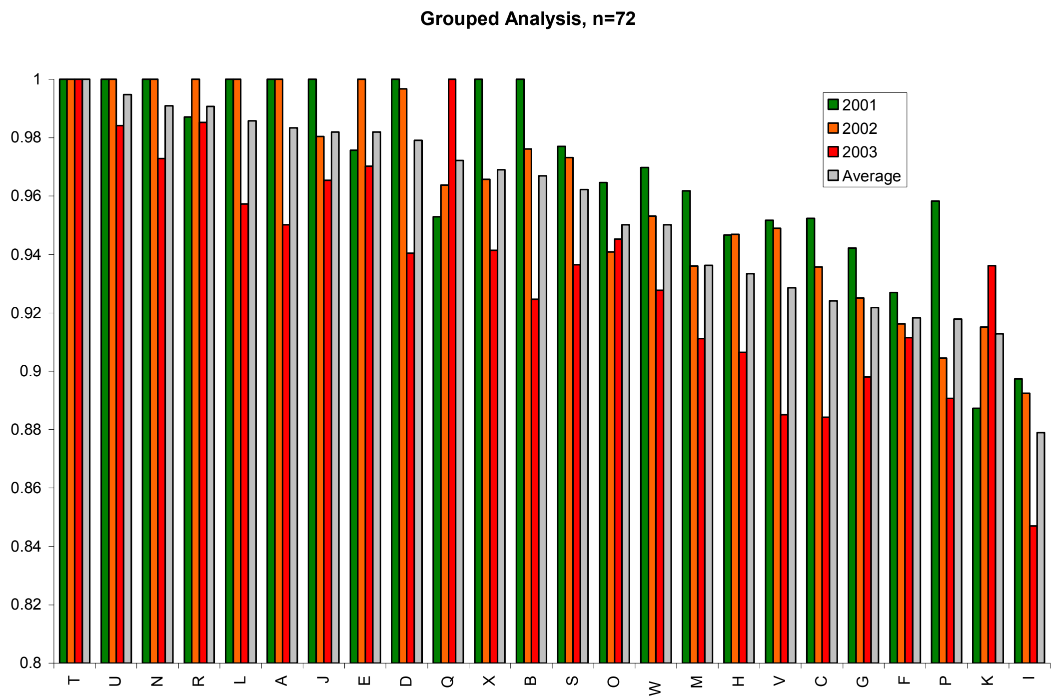


Figure 23. Grouped Analysis, 72 Observations Considered as Separate DMUs.

The results of the grouped analysis of the DMUs are depicted above. The results of both the grouped and ungrouped efficiency analyses are largely congruent, although expanding the analysis to 72 DMUs allows for more differentiation. It is interesting to note that, based on the mean efficiency scores, hospitals P, K, and I occupy the bottom positions, while hospital T is at the top in both analyses (see Figure 24.)

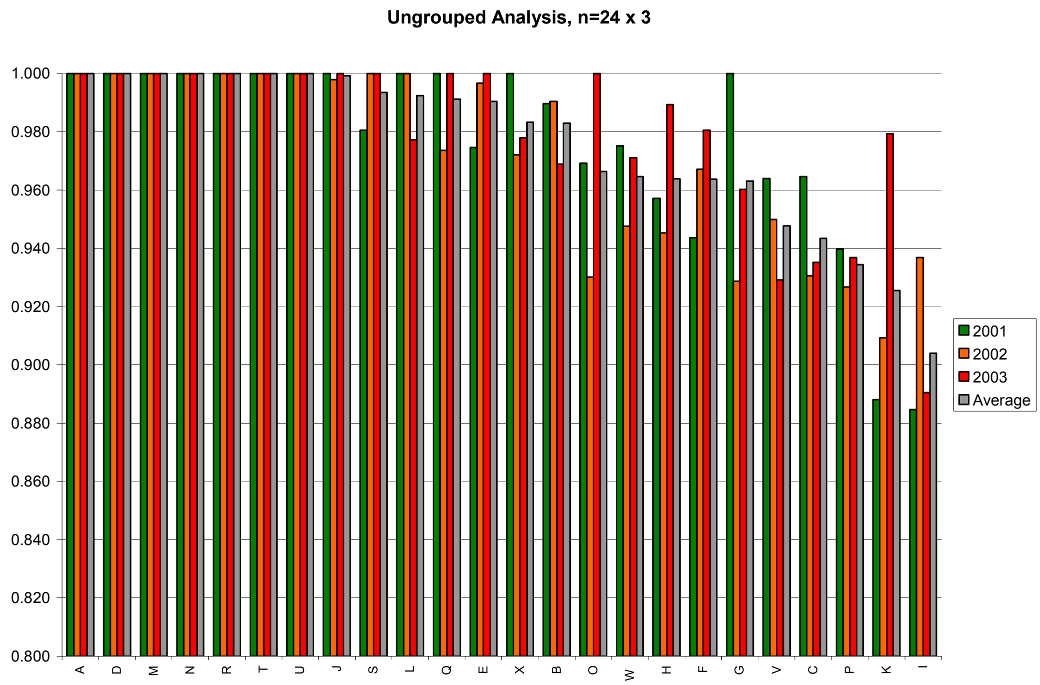


Figure 24. Ungrouped Analysis, 24 DMUs * Three Years.

The ungrouped efficiency analyses for all three years is similar to the grouped analysis.

Slacks from the 2001-2003 output and input oriented analyses suggest that there may be excess costs in the system. Focusing on the input orientation, one notices quite readily that the same low performing facilities are again present each year. Specifically, facilities K, I, F, and P appear to have cost and access issues primarily. By this analysis, $41.06 million might have been reallocated based solely on slacks from the efficiency analysis. This $41 million represents less than one-half of one percent of the $12 billion expended over these three years. The following tables (Tables 32-34) provide the inefficient facilities grouped by year and sorted by technical efficiency score. One will notice that access and workload slacks are quite nominal, while quality slacks show facilities with significant shortages; however, no facility exhibited a shortage for all three years of the analysis. It is of some concern that the quality / prevention slack for the year 2003 is in excess of what was experienced in 2002 and 2001.

Table 32. Inefficiencies, 2001.

DMU	YR	Score	C Slack	A Slack	Q Slack	W Slack	Inefficiency ($)
K	2001	0.887	(0.481)	-	-	-	$ 0.62
I	2001	0.897	(0.481)	0.081	-	-	$ 0.62
F	2001	0.927	(0.371)	0.038	-	-	$ 0.69
G	2001	0.942	(0.317)	-	0.633	-	$ 0.73
H	2001	0.947	(0.254)	-	-	-	$ 0.78
V	2001	0.952	(0.248)	-	-	-	$ 0.78
C	2001	0.952	(0.234)	0.019	-	-	$ 0.79
Q	2001	0.953	(0.210)	-	-	-	$ 0.81
P	2001	0.958	(0.175)	0.037	-	-	$ 0.84
M	2001	0.962	(0.150)	-	-	0.046	$ 0.86
O	2001	0.965	(0.167)	-	-	-	$ 0.85
W	2001	0.970	(0.138)	0.126	-	-	$ 0.87
E	2001	0.976	(0.132)	0.020	-	-	$ 0.88
S	2001	0.977	(0.134)	0.007	1.000	-	$ 0.87
R	2001	0.987	(0.057)	-	0.173	-	$ 0.94

Table 33. Inefficiencies, 2002.

DMU	YR	Score	C Slack	A Slack	Q Slack	W Slack	Inefficiency ($)
I	2002	0.892	(0.506)	-	-	-	$ 0.60
P	2002	0.905	(0.406)	0.064	-	-	$ 0.67
K	2002	0.915	(0.362)	0.098	-	0.031	$ 0.70
F	2002	0.916	(0.433)	-	-	-	$ 0.65
G	2002	0.925	(0.411)	0.016	-	-	$ 0.66
C	2002	0.936	(0.320)	-	-	-	$ 0.73
M	2002	0.936	(0.250)	-	0.219	-	$ 0.78
O	2002	0.941	(0.284)	0.010	-	-	$ 0.75
H	2002	0.947	(0.255)	-	-	-	$ 0.77
V	2002	0.949	(0.263)	0.074	-	-	$ 0.77
W	2002	0.953	(0.218)	0.016	-	-	$ 0.80
Q	2002	0.964	(0.162)	-	1.181	-	$ 0.85
X	2002	0.966	(0.195)	0.047	-	-	$ 0.82
S	2002	0.973	(0.156)	-	3.430	-	$ 0.86
B	2002	0.976	(0.100)	-	-	-	$ 0.90
J	2002	0.980	(0.080)	0.033	-	-	$ 0.92
D	2002	0.997	(0.019)	0.019	-	-	$ 0.98

Table 34. Inefficiencies, 2003.

DMU	YR	Score	C Slack	A Slack	Q Slack	W Slack	Inefficiency ($)
I	2003	0.847	(0.745)	-	-	-	$ 0.98
C	2003	0.884	(0.591)	0.104	0.103	-	$ 0.47
V	2003	0.885	(0.603)	0.113	-	-	$ 0.55
P	2003	0.891	(0.467)	0.056	1.000	-	$ 0.55
G	2003	0.898	(0.565)	-	-	-	$ 0.63
H	2003	0.907	(0.452)	-	0.263	-	$ 0.57
M	2003	0.911	(0.352)	-	-	0.090	$ 0.64
F	2003	0.912	(0.454)	-	0.618	-	$ 0.70
B	2003	0.925	(0.323)	-	3.000	-	$ 0.64
W	2003	0.928	(0.342)	0.048	-	-	$ 0.72
K	2003	0.936	(0.254)	-	1.000	-	$ 0.71
S	2003	0.936	(0.377)	-	-	-	$ 0.78
D	2003	0.940	(0.358)	-	1.445	-	$ 0.69
X	2003	0.941	(0.338)	-	0.064	-	$ 0.70
O	2003	0.945	(0.266)	-	0.853	-	$ 0.71
A	2003	0.950	(0.201)	-	1.000	-	$ 0.77
L	2003	0.957	(0.195)	-	-	-	$ 0.82
J	2003	0.965	(0.145)	-	0.781	-	$ 0.82
E	2003	0.970	(0.159)	-	-	-	$ 0.87
N	2003	0.973	(0.157)	0.006	1.000	-	$ 0.85
U	2003	0.984	(0.058)	-	0.030	0.082	$ 0.85
R	2003	0.985	(0.067)	-	-	-	$ 0.94

Table 35 provides the input oriented efficiency projections to the frontier. The percentages provided suggest that many of the lower performing facilities identified in the analysis have excess funds and are not providing enough output in comparison to the referent set. For example, Facility I in 2003 may have received 15.3% more funding than required to produce the same workload, quality, and access outputs. Also, consider Facility I in 2001. Not only did the hospital receive 10.26% more funding than required, it also experienced an access shortage as represented by the 1.96% projection.

Table 35. Projections, 2001-2003.

The projections from the input-oriented BCC analysis suggest that certain facilities may be overfunded.

	Cost 01	Work 01	Access 01	Quality 01	Cost 02	Work 02	Access 02	Quality 02	Cost 03	Work 03	Access 03	Quality 03
K	-11.27%	0.00%	0.00%	0.00%	-8.49%	2.24%	2.36%	0.00%	-6.39%	0.00%	0.00%	33.33%
I	-10.26%	0.00%	1.96%	0.00%	-10.75%	0.00%	0.00%	0.00%	-15.30%	0.00%	0.00%	0.00%
F	-7.30%	0.00%	0.92%	0.00%	-8.39%	0.00%	0.00%	0.00%	-8.84%	0.00%	0.00%	20.60%
G	-5.78%	0.00%	0.00%	15.82%	-7.50%	0.00%	0.38%	0.00%	-10.19%	0.00%	0.00%	0.00%
H	-5.34%	0.00%	0.00%	0.00%	-5.31%	0.00%	0.00%	0.00%	-9.35%	0.00%	0.00%	8.76%
V	-4.82%	0.00%	0.00%	0.00%	-5.11%	0.00%	1.82%	0.00%	-11.49%	0.00%	2.80%	0.00%
C	-4.77%	0.00%	0.46%	0.00%	-6.43%	0.00%	0.00%	0.00%	-11.58%	0.00%	2.58%	3.44%
Q	-4.71%	0.00%	0.00%	0.00%	-3.63%	0.00%	0.00%	39.36%	0.00%	0.00%	0.00%	0.00%
P	-4.17%	0.00%	0.88%	0.00%	-9.54%	0.00%	1.55%	0.00%	-10.93%	0.00%	1.37%	33.33%
M	-3.83%	3.58%	0.00%	0.00%	-6.40%	0.00%	0.00%	7.30%	-8.88%	8.66%	0.00%	0.00%
O	-3.53%	0.00%	0.00%	0.00%	-5.92%	0.00%	0.24%	0.00%	-5.48%	0.00%	0.00%	28.43%
W	-3.02%	0.00%	3.09%	0.00%	-4.69%	0.00%	0.39%	0.00%	-7.22%	0.00%	1.16%	0.00%
E	-2.43%	0.00%	0.49%	0.00%	0.00%	0.00%	0.00%	0.00%	-2.99%	0.00%	0.00%	0.00%
S	-2.30%	0.00%	0.17%	33.33%	-2.68%	0.00%	0.00%	342.98%	-6.36%	0.00%	0.00%	0.00%
R	-1.30%	0.00%	0.00%	5.75%	0.00%	0.00%	0.00%	0.00%	-1.48%	0.00%	0.00%	0.00%
X	0.00%	0.00%	0.00%	0.00%	-3.42%	0.00%	1.15%	0.00%	-5.87%	0.00%	0.00%	1.60%
B	0.00%	0.00%	0.00%	0.00%	-2.39%	0.00%	0.00%	0.00%	-7.54%	0.00%	0.00%	300.00%
J	0.00%	0.00%	0.00%	0.00%	-1.95%	0.00%	0.80%	0.00%	-3.45%	0.00%	0.00%	26.03%
D	0.00%	0.00%	0.00%	0.00%	-0.32%	0.00%	0.45%	0.00%	-5.96%	0.00%	0.00%	48.18%
T	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
U	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	-1.59%	7.99%	0.00%	0.99%
N	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	-2.71%	0.00%	0.15%	33.33%
L	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	-4.27%	0.00%	0.00%	0.00%
A	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	-4.98%	0.00%	0.00%	33.33%

In order to evaluate the slack findings, a RAM model (grouped) for 2001-2003 was run. The results of the RAM model as shown in Table 36 imply that more than $18 million might have been reallocated to improve efficiency for the years 2001-2003. Facilities K and I are again identified in all years as inefficient due to costs. Looking at Facility I, one can again see that it is overfunded in 2001 by $1.2 million, in 2002 by $1.05 million, and in 2003 by $1.1 million. Most interestingly, the overall cost slack has increased from 2001 to 2003 from roughly $2.5 million to $11 million.

Table 36. RAM Results, 2001-2003.

The slacks from the RAM model sorted by year are provided.

Year	DMU	C slack	W slack	Q slack	A slack	Inefficiency ($)
2001	I	0.179	0.098	1.62	0.046	$ 1.20
2001	K	0.282		1.178		$ 1.33
2001	O			0.355		
2001	V			0.5		
2001	R			0.959		
2001	B			0.962		
2001	E		0.091	1.016		
2001	S			1.03		
2001	C			1.068		
2001	H			1.136		
2002	I	0.052	0.144	0.57		$ 1.05
2002	X	0.064		0.015	0.064	$ 1.07
2002	G	0.149		1.515	0.018	$ 1.16
2002	K	0.183		1.797	0.039	$ 1.20
2002	D			0.148	0.006	
2002	A			0.38		
2002	P		0.191	0.429	0.078	
2002	W			0.646		
2002	J		0.07	1.006	0.039	
2002	H			1.282		
2002	O			1.318		
2002	V			1.479	0.078	
2002	B			1.859		
2002	C			2.137		
2003	U	0.005	0.03		0.04	$ 1.01
2003	M	0.013	0.121	1.979	0.018	$ 1.01
2003	L	0.034		2.289		$ 1.03
2003	K	0.035		2.473		$ 1.04
2003	I	0.094	0.323	2.487		$ 1.10
2003	A	0.125		1.463		$ 1.13
2003	S	0.153	0.049	1.798		$ 1.17
2003	D	0.16		1.158		$ 1.17
2003	V	0.29		2.55	0.1	$ 1.34
2003	G	0.301		1.311		$ 1.35
2003	O			0.686		
2003	P		0.186	1.217	0.071	
2003	H		0.036	1.352		
2003	W		0.039	1.833	0.003	
2003	B			2.914		
2003	C		0.303	4.23	0.081	

An analysis of referent sets indicates that neither the MEDCENS nor the ACHs are over-represented as referent entities; however, the primary referent entity based upon frequency was a MEDDAC in 2002 (R02). A separate analysis of MEDDACS and MEDCEN subgroupings was therefore deemed necessary to ensure that the previous analyses were representative. Table 37 provides the summary of the referent units for the BCC-I analysis of 2001-2003.

Table 37. Referent Sets, 2001-2003.

The input oriented referent sets are reflected in the following table. The facility R02 is overrepresented in the analysis.

DMU	A01	B01	D01	J01	L01	N01	T01	U01	X01	A02	E02	L02	N02	R02	T02	U02
A01	x															
B01		x														
C01					x						x	x				
D01			x													
E01									x		x			x		
F01									x		x			x		
G01					x	x					x			x		
H01					x	x					x			x		
I01		x			x									x		
J01				x												
K01	x			x				x								
L01					x											
M01	x							x								
N01						x										
O01	x				x	x								x		
P01	x	x														
Q01	x				x	x								x		
R01									x				x			
S01									x				x			
T01							x									
U01								x								
V01					x	x					x			x		
W 01		x			x								x			
X01									x							

(continued on next page)

DMU	A01	B01	D01	J01	L01	N01	T01	U01	X01	A02	E02	L02	N02	R02	T02	U02
A02										x						
B02	x			x	x									x		
C02					x	x					x			x		
D02			x				x									
E02											x					
F02						x			x		x			x		
G02									x		x			x		
H02					x	x					x			x		
I02	x			x	x									x		
J02				x				x								
K02										x						
L02												x				
M02	x							x								x
N02													x			
O02					x						x			x		
P02				x				x								
Q02	x					x								x		
R02														x		
S02						x			x				x			
T02															x	
U02																x
V02									x		x			x		
W 02		x			x									x		
X02									x				x			
A03	x			x				x								
B03	x			x				x								
C03									x					x		
D03						x			x				x			
E03						x			x		x		x			
F03	x					x					x			x		
G03						x			x		x			x		
H03	x					x								x		
I03	x			x	x									x		
J03	x			x										x		
K03	x			x				x								
L03	x				x	x								x		
M03	x							x								x
N03	x					x								x		
O03	x					x								x		
P03				x				x								
Q03																
R03					x	x								x		
S03									x		x			x		
T03																
U03								x								x
V03					x						x			x		
W 03		x			x									x		
X03					x				x					x		

MEDCEN and Army Community Hospital Subgroup Analysis

An inherent assumption in comparing the facilities above is that they evaluated in reference to the performance of efficient DMUs in a peer group. Specifically, the removal of GME and readiness costs is assumed to have allowed comparisons among the hospitals despite differences in size. This assumption requires some additional investigation. Individual analysis for the MEDCENs and ACHs was performed to address these concerns.

The results of BCC-I analysis on the ACHs are congruent with the original findings. The ordering is nearly identical, and the correlation coefficient is .90. While not perfect, the simplifying assumption that subtracting GME expenditures accounts for much of the MEDCEN mission appears to work well from the perspective of the ACHs. In Figure 25, the analysis of hospitals without the inclusion of MEDCENs results in the identification of many of the same inefficient DMUs as in the analysis that included MEDCENs (e.g., P, K, and I). Further, the efficiency scores for the hospitals generated by the analysis with MEDCENs and the analysis without them is correlated ($p<.01$) at the .90 level.

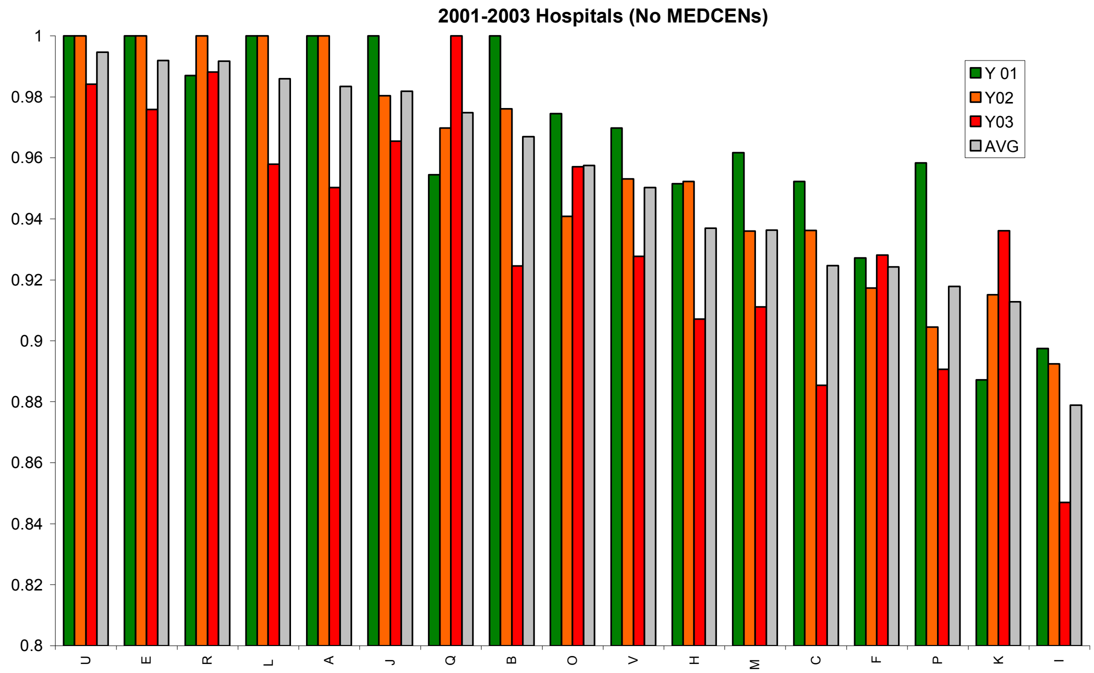


Figure 25. Hospitals Only BCC Analysis, 2001-2003.

The analysis of efficiency without MEDCENs is largely congruent with original findings. It is easy to notice that the bottom and top efficiency scores have not changed from the grouped analysis. The correlation between the average efficiency scores for the grouped and ungrouped analyses is .90.

The results are not quite as nice for the MEDCENs when analyzed by themselves. From the next diagram (Figure 26), one immediately notes that there is disparity in identifying the lower performing group. The correlation between the average efficiency score for 2001-2003 for the MEDCENs as part of the original set of DMUs and as calculated as a separate peer group is only .389, hardly impressive.

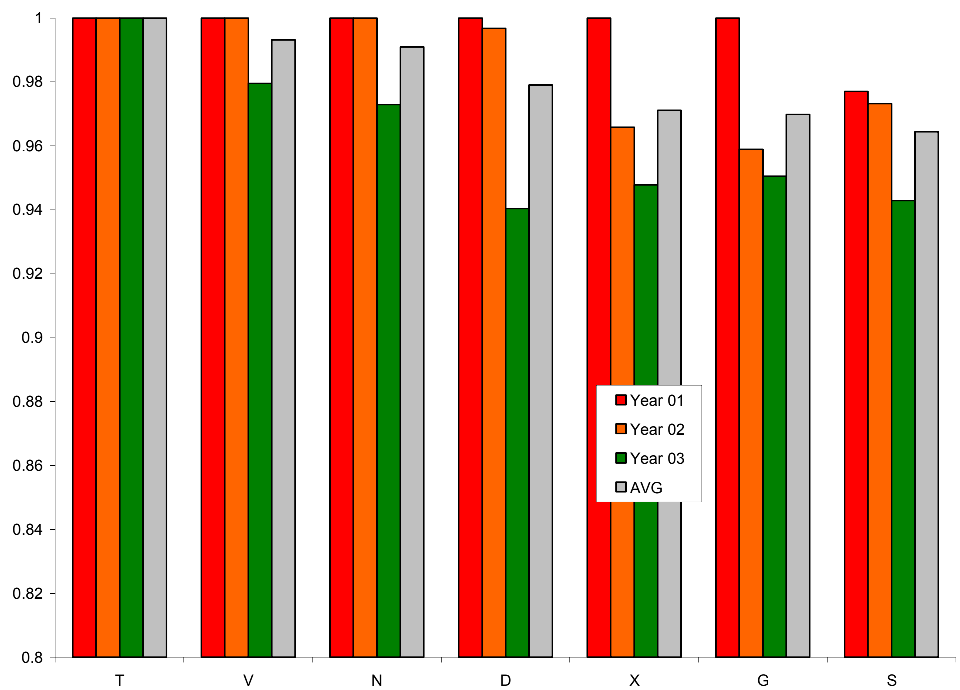


Figure 26. MEDCEN Only BCC Analysis, 2001-2003.
Efficiency analysis for the MEDCEN facilities indicates that separate analysis of these DMUs may be useful.

Slacks generated by the ungrouped analysis provide additional management insight into cost, quality, and access shortages. Table 38 illustrates the slacks for MEDCEN facilities for the BCC-I analysis. It is interesting to note that these slacks are similar to those estimated in the grouped analysis. The potential reallocation of the cost slacks for MEDCENS only would represent $8 million.

Table 38. MEDCEN Only BCC Results, 2001-2003.

DMU	Score	C Slack	A Slack	Q Slack	W Slack	Trans. $
S01	0.977029	(0.1337)	0.0070	1.0000	-	0.874852
G02	0.958944	(0.2250)	-	0.3291	-	0.798516
X02	0.965816	(0.1946)	0.0468	-	-	0.823164
S02	0.973231	(0.1564)	-	3.4298	-	0.855217
D02	0.996777	(0.0191)	0.0190	-	-	0.981081
V03	0.979506	(0.1075)	0.1249	1.0000	0.1644	0.898077
D03	0.940367	(0.3577)	-	1.4454	-	0.699283
S03	0.942975	(0.3386)	0.0141	-	-	0.712767
X03	0.947813	(0.3007)	-	0.2424	-	0.7403
G03	0.95056	(0.2740)	-	0.7903	-	0.760332
N03	0.972938	(0.1575)	0.0063	1.0000	-	0.854277

Similarly, the slacks for the RAM model of the MEDCENS identified nearly the same inefficient facilities each year. The slack values, however, are different. This difference is, in part, due to the use of 1-θ* as the slack driver for the BCC-I models (for θ*<1). Table 39 provides the results.

Table 39. MEDCEN Only RAM results, 2001-2003.

The RAM model results follow.

DMU	C slack	W slack	Q slack	A slack	Inefficiency ($)
S01			1.03		
X02	0.064		0.015	0.064	$ 1.07
D02			0.148	0.006	
V02			0.928	0.055	
G02		0.116	1.199	0.031	
D03	0.097		1.219		$ 1.10
S03	0.153	0.049	1.798		$ 1.17
G03		0.168	1.132		
V03		0.219	2.058	0.12	

Table 40 provides the slacks obtained from the input-oriented ACH analysis. A comparison of the findings of this analysis and the grouped analysis suggests that there is widespread congruence. The cost adjustments implied by the slack values are nearly $30 million over the three-year period.

Table 40. ACH Only BCC Results 2001-2003.

The slacks from the ungrouped analysis of the ACHs are largely identical to those from the analysis of the MEDCENs.

DMU	Year	Score	C Slack	A Slack	Q Slack	W Slack	Transl. $
K	2001	0.887261	(0.481)	-	-	-	$ 0.62
I	2001	0.89742	(0.481)	0.081	-	-	$ 0.62
F	2001	0.927182	(0.370)	0.039	0.032	-	$ 0.69
H	2001	0.951539	(0.231)	-	0.431	-	$ 0.79
C	2001	0.952337	(0.234)	0.019	-	-	$ 0.79
Q	2001	0.954541	(0.202)	-	0.972	-	$ 0.82
P	2001	0.95829	(0.175)	0.037	-	-	$ 0.84
M	2001	0.961718	(0.150)	-	-	0.046	$ 0.86
W	2001	0.969798	(0.138)	0.126	-	-	$ 0.87
O	2001	0.974525	(0.121)	-	1.644	-	$ 0.89
R	2001	0.987034	(0.057)	-	0.173	-	$ 0.94
I	2002	0.892492	(0.506)	-	-	-	$ 0.60
P	2002	0.904585	(0.406)	0.064	-	-	$ 0.67
K	2002	0.915142	(0.362)	0.098	-	0.031	$ 0.70
F	2002	0.917373	(0.426)	-	0.156	-	$ 0.65
M	2002	0.936049	(0.250)	-	0.219	-	$ 0.78
C	2002	0.936259	(0.317)	-	0.050	-	$ 0.73
O	2002	0.940837	(0.284)	0.010	-	-	$ 0.75
H	2002	0.95233	(0.229)	-	0.484	-	$ 0.80
W	2002	0.953102	(0.218)	0.016	-	-	$ 0.80
Q	2002	0.969856	(0.135)	-	1.840	-	$ 0.87
B	2002	0.976141	(0.100)	-	-	-	$ 0.90
J	2002	0.980477	(0.080)	0.033	-	-	$ 0.92
I	2003	0.846952	(0.745)	-	-	-	$ 0.47
C	2003	0.885405	(0.584)	0.110	0.390	-	$ 0.56
P	2003	0.890699	(0.467)	0.056	1.000	-	$ 0.63
H	2003	0.907088	(0.449)	-	0.646	-	$ 0.64
M	2003	0.911178	(0.352)	-	-	0.090	$ 0.70
B	2003	0.924569	(0.323)	-	3.000	-	$ 0.72
W	2003	0.927787	(0.342)	0.048	-	-	$ 0.71
F	2003	0.928144	(0.369)	-	2.513	-	$ 0.69
K	2003	0.936127	(0.254)	-	1.000	-	$ 0.78
A	2003	0.950234	(0.201)	-	1.000	-	$ 0.82
O	2003	0.957068	(0.209)	-	2.899	-	$ 0.81
L	2003	0.957932	(0.192)	-	0.376	-	$ 0.83
J	2003	0.965468	(0.145)	-	0.781	-	$ 0.86
E	2003	0.9759	(0.128)	0.004	1.666	-	$ 0.88
U	2003	0.984119	(0.058)	-	0.030	0.082	$ 0.94
R	2003	0.988218	(0.053)	-	1.786	-	$ 0.95

The RAM models run solely on the Army Community Hospitals produced results similar to the BCC model as shown in Table 41. Once again, facilities K and I are identified as inefficient in for all years and in all models.

Table 41. ACH Only RAM Results, 2001-2003.

The slacks for the ACH RAM model are shown below.

DMU	YEAR	C slack	W slack	Q slack	A slack	Dollars
I	2001	0.197		1.558	0.051	$ 1.22
K	2001	0.326		1.053	0.009	$ 1.39
R	2001			0.761	0.006	
C	2001			0.846	0.006	
B	2001			0.931	0.018	
H	2001			0.963		
I	2002	0.131		0.517		$ 1.14
K	2002	0.246		1.56	0.06	$ 1.28
A	2002			0.38		
P	2002		0.256	0.676		
H	2002			1.067		
O	2002			1.114		
J	2002		0.01	1.129		
B	2002			1.982		
M	2003	0.018	0.111	1.977	0.019	$ 1.02
K	2003	0.035		2.473		$ 1.04
H	2003	0.04		1.237		$ 1.04
U	2003	0.042	0.019	0.069		$ 1.04
L	2003	0.047		2.216		$ 1.05
A	2003	0.125		1.463		$ 1.13
B	2003	0.166		4.145	0.084	$ 1.18
I	2003	0.314	0.021	2.481		$ 1.37
O	2003			0.301		
W	2003			0.646		
P	2003		0.245	1.442		
W	2003		0.042	1.842		

Replacement of Contrived Population Variable with Enrollment Only, 2001-2003

To be thorough in the investigation of input and output specification, the population variable, which was composed of elements of enrollment and population, was replaced with pure, untransformed enrollment values for the 2001-2003 years. Theoretically, the factor score should contain nearly all of the information that represents enrollment due to the high shared variance discussed previously. Further, due to the units invariant property of DEA formulations, the use of raw enrollment numbers should change the results very little. The findings are remarkable. The scores and slacks were identical for all DMUs, indicating that there was not enough movement in the substitution to shift the results.

However, splitting the RVUs and RWPs into separate outputs and inserting them as raw data did change the results somewhat. The correlation between facility rankings of the original and modified analyses is .894 ($p<.01$), which suggests that differences exist when outputs are split. This finding is not new in the literatures. Magunsson (1993) illustrated how one simple change in the output measurement could result in large discrepancies in efficiency scores. By using a single composite metric, which encapsulates the workload factors for the facility, this pitfall might be avoided.

As another check of the stability of the efficiency rankings, the quality and access variables were removed from the analysis, and the BCC-I model was re-run. The results are again correlated (ρ=.873, $p<.01$), although there certainly are some differences.

A final check of the stability of the transformed data versus the complete set of untransformed data (as specified in Table 9) was performed for the three-year panel series. The correlation is less impressive (.758, $p<.01$) but still reasonable. Since efficiency scores will be used in additional cost models as fixed regressor terms, the differences in measurement should be less severe. Specifically, the correlation of the DEA efficiency measures suggests that the selection of any one of the input and output variations will provide similar (albeit not identical) results in loglinear models. Further, robust regression assumes that there exist errors in variables, so slight measurement error is less problematic.

Summary and Conclusions for Efficiency Analyses

In the past chapter, DMU efficiency has been analyzed in several different ways. Efficiency for year 2001 was analyzed separately using DEA, SFA, and COLS analyses and this analysis was then extended to the years 2001 and 2002. All three years of data were then evaluated and then separate analyses for MEDCENS and ACHs were performed. The COLS estimator provided less reasonable results than did the DEA and SFA estimators. While the efficient DMUs differed by method and by year, Facilities K and I were routinely ranked near the bottom of all analyses independent of time, and Facilities T and U were near the top.

From this analysis, a manager is able to draw conclusions regarding consistently low and consistently high performing DMUs; however, it is not possible to determine exactly why without further analysis. From a managerial perspective, ACHs K and I are ranked low in efficiency. These facilities are geographically diverse and have no readily identifiable similarities that were not taken into consideration. An analysis of their referent groups suggests that Facility R is referent for Facility F and that Facilities A, J, and U are referent for Facility I. The referent facilities are geographically diverse hospitals with no obvious similarities. High performing entities vary quite a bit more by year; however, for MEDCENs, facility T remains Pareto-Koopmans efficient in all analyses. ACH facilities U, E, L, and A have fairly high efficiency scores in all analyses. It is possible that omitted variables explain the deviations, but, if the efficiency model is sufficiently complete, then help in identifying the reasons for the efficiency is found in the cost, quality, and access projections.

It is prudent to investigate cost models and test the forecasting ability of these models prior to the implementation of any change based solely on the above models. The next chapter will evaluate multiple models that include efficiency as a predictor term or random variable. For further analysis of cost, the DEA input orientation is selected as it appears to be more congruent with the other efficiency analysis and the output orientation is abandoned. Both SFA formulations are initially selected; however, the truncated normal model is later abandoned in favor of the half-normal formulation.

Chapter 6: Cost Models for Military Hospitals

Cross-Sectional Analysis, 2001

With the results of the efficiency estimates completed, an analysis of cost models that include these estimates is possible to obtain quantitative estimates of cost by changing the values of the regressor variables. The SFA, OLS, and robust and ridge regression models are discussed in this chapter.

<u>Stochastic Frontier Models</u>

The SFA models (both half-normal and truncated normal) produced both the efficiency estimates as discussed previously and other parameter estimates relevant to the cost models. The models evaluated were the half-normal and truncated normal formulations as shown. For convenience of the reader, the models and their variables that were provided in Chapter 4, equations (62) and (63), are reproduced. As in the linear models with DEA efficiency scores, a generalized loglinear (Cobb-Douglas) formulation with a single interaction term associated with workload is specified.[8]

$$\begin{gathered}
\text{SFA 1: HN (Half Normal)} \\
C = [1 \;\; W \;\; P \;\; Q \;\; A \;\; M \;\; M{*}W][\beta] + v + u \\
\varepsilon = u + v \text{ (error and random noise)} \\
v \sim N(0, \sigma_v^2) \text{ (random noise)} \\
u \sim N^{+}(0, \sigma_u^2)
\end{gathered}$$

u is inefficiency.

[8] A translog formulation was considered; however, the previous variable transformations suggested that the simple log-linear model would be effective.

$$\text{SFA 2: TN (Truncated Normal)}$$
$$C = [1\ W\ \ P\ \ Q\ \ A\ \ M\ \ M{*}W][\beta] + v + u$$
$$\varepsilon = u + v \text{ (error and random noise)}$$
$$v \sim N(0, \sigma_v^2) \text{ (random noise)}$$
$$u \sim N^+(\mu, \sigma_u^2)$$

u is inefficiency.

Recall that *W* is workload, *P* is population, *Q* is a quality / prevention proxy, *A* is an access proxy, *M* is a MEDCEN indicator variable and *M*W* is the interaction between *M* and *W*.

As shown in Tables 42 and 43, only one variable was strongly related to cost in both model formulations (the half-normal and the truncated normal), and that variable was workload. In Table 42, the half-normal formulation, one can see that the model is not statistically significant by evaluation of the log likelihood function (which is to be expected for the small number of observations) and that the coefficient for workload is .99, implying that there is nearly one to one effect of the workload variable on the cost variable. The directionality is positive, which matches the expectation. Further, the standard error for the workload term is quite large (.21). In Table 43, the SFA Truncated Normal formulation, one can see that the model is insignificant and that the workload coefficient is also near unity (1.09) and has a significant error terms as well (.20). It is also unfortunate that the coefficients have relatively large standard errors, as this fact implies that there will be a lack of stability in future forecasting models.

Table 42. SFA Half Normal Coefficients, 2001.

The SFA half-normal model is not statistically significant ($p<.632$); however, at .945, the approximated R^2 indicates a strong relationship between the model and the data. The MLE coefficient estimates for the SFA half-normal analysis for year 2001 reveals only one statistically significant parameter (workload) other than the constant. Negative numbers are in parenthesis

Number of observations	24
Iterations completed	101
Log likelihood function	6.310
Chi-Square Probability	0.632
R^2	.945

Standardized Variables	**Coefficient**	**SE**	**P[Z]>z**
Constant	(0.29)	0.10	<.001
P	(0.09)	0.13	0.49
W	0.99	0.21	<.001
Q	(0.01)	0.04	0.80
A	0.02	0.05	0.69
M	0.14	0.33	0.68
M*W	(0.01)	0.36	0.98
σ_v	.00		
σ_u	.14		

Table 43. SFA Truncated Normal Coefficients, 2001.

Once again, the model is not statistically significant (p<.459), but the R^2 is reasonable (.937). The coefficient estimates for the SFA Truncated Normal analysis are similar to those estimated in the Half-Normal analysis.

Number of observations	24
Iterations completed	101
Log likelihood function	6.936
Chi-Square Probability	0.459
R^2	.937

Standardized Variables	**Coefficient**	**SE**	**P[Z]>z**
Constant	(0.28)	0.07	<0.001
P	(0.20)	0.16	0.20
W	1.09	0.20	<.001
Q	(0.01)	0.04	0.75
A	0.01	0.10	0.91
M	0.20	0.36	0.58
M*W	(0.06)	0.40	0.88
μ_u	2.48	20.20	
σ_v	0.00		
σ_u	0.83364		

It is interesting to note that the primary coefficient estimates for all variables other than the log of the workload variable are near zero. The implication from the SFA analysis is that cost is largely driven by workload. The only statistically significant variables are the constant and workload under both formulations.

The SFA models produce estimates for cost, which might be used for resource allocation. The estimates obtained from both of the SFA models are highly correlated as shown in Table 44 and significantly different from zero. The ability of the estimates produced by SFA to be used in forecasting models is yet to be discussed.

Table 44. Correlation between SFA Models, 2001.

The correlation for the SFA-HN / TN estimates and the response variable is high (.972 and .968 respectively), which roughly indicates that about 94.5% of the variance of the budget is accounted for by the SFA model. Five percent, however, equates to millions of dollars. Other models will be investigated for improvement.

	1. Observed	2. HN SFA	3. TN SFA
1. Observed	1.000		
2. HN SFA	0.972**	1.000	
3. TN SFA	0.968**	0.998**	1.000

**$p<.001$

Linear Models with Efficiency Scores

The first loglinear model investigated is based on the traditional Cobb-Douglas cost function just as in the SFA models. Using Pareto-Koopmans efficiency as a binary (dummy) variable and evaluating the interactions as discussed previously results in the following model, which is a replication of (65) from Chapter 4.

$$\begin{gathered}
\text{Linear 1:}\\
C=[1\;\; W\; P\; Q\; A\;\; M\;\; M{*}W][\beta]+[E\;\; E{*}W\; E{*}P\; E{*}Q\; E{*}A][\gamma]+\varepsilon\\
E\equiv\text{Pareto Koopmans Efficiency Indicator},\\
\text{where } E \text{ is the vector composed of } E_i=\begin{cases}0 \text{ if DMU}_\text{i}\text{ is not efficient}\\ 1 \text{ if DMU}_\text{i}\text{ is efficient}\end{cases}\text{and}\\
M \text{ is the vector composed of } M_i=\begin{cases}0 \text{ if DMU}_\text{i}\text{ is not a MEDCEN}\\ 1 \text{ if DMU}_\text{i}\text{ is a MEDCEN}\end{cases}
\end{gathered}$$

In this model, the equation for inefficient units ($E = 0$) simplifies to $C=\beta_0+\beta_1 W+\beta_2 P+\beta_3 Q+\beta_4 A+\beta_5 M+\beta_6 M*W+\varepsilon$, while the equation for efficient units ($E=1$) is given by the following.

$$C = \beta_0 + \gamma_0 + (\beta_1 + \gamma_1)W + (\beta_2 + \gamma_2)P + (\beta_3 + \gamma_3)Q + (\beta_4 + \gamma_4)A + (\beta_5 + \gamma_5)M + (\beta_6 + \gamma_6)M * W + \varepsilon$$

For further information, the interested reader is directed to Bardhan, Cooper and Kumbhakar (1998).

Ordinary Least Squares (OLS) is used to estimate the model. The results of the regression are provided in Table 45. In the case of this linear model, caution is advised in the interpretation of the inferential results, as the sample size is small in relation to the number of parameters. Nevertheless, the model appears to be a strong fit as indicated by the significance ($F(11,12)$=43.2576, $p<.001$) and the R^2 of .9754 (unadjusted, .9529 adjusted). Further, two coefficients are statistically significant at the .05 level: Workload and Efficiency. The directionality of both variables is in accordance with economic theory in that an increase in workload should drive up costs, while an increase in efficiency should reduce costs. By evaluating the magnitude of the coefficients, one notes that, while workload is the largest driver of costs in this model, the efficiency term is still a significant contributor (W=.7581 versus E=-.1623).

Table 45. Linear Model 1 Regression Results, 2001.

The parameter estimates for the year 2001 linear model with Pareto-Koopmans efficiency included follow. The estimates are least squares solutions. Because of the small sample size relative to the number of parameters, the results need to be interpreted with caution. The adjusted R^2 is 95.3%, almost a one percent increase over the SFA model. The interesting estimates are for the parameters associated with workload and efficiency.

Summary measures

Multiple R	0.9876
R-Square	0.9754
Adj R-Square	0.9529
StErr of Est	0.2171

ANOVA Table

Source	df	SS	MS	F	p-value
Explained	11	22.4342	2.0395	43.2576	0.0000
Unexplained	12	0.5658	0.0471		

Regression coefficients

	Coefficient	Std Err	t-value	p-value	Lower Limit (5%)	Upper limit (95%)
Constant	0.0000	0.0443	0.0000	1.0000	-0.0966	0.0966
M	0.0417	0.1967	0.2120	0.8356	-0.3870	0.4704
MW	0.1719	0.2180	0.7885	0.4457	-0.3031	0.6469
P	-0.0135	0.1220	-0.1109	0.9135	-0.2793	0.2523
W**	0.7581	0.1819	4.1676	**0.0013**	0.3618	1.1544
Q	0.0916	0.0729	1.2572	0.2326	-0.0672	0.2504
A	0.0299	0.0614	0.4879	0.6344	-0.1037	0.1636
E*	-0.1623	0.0737	-2.2028	**0.0479**	-0.3228	-0.0018
EP	-0.1001	0.1169	-0.8562	0.4087	-0.3547	0.1546
EW	0.1112	0.1162	0.9570	0.3574	-0.1420	0.3645
EQ	0.0528	0.0625	0.8449	0.4147	-0.0834	0.1890
EA	0.0581	0.0564	1.0299	0.3234	-0.0648	0.1811

*$p<.05$, **$p<.01$

The managerial significance of the difference between the SFA and linear models appears to be nominal until one considers the impact in terms of dollars. The difference between accounting for 94.5% and 95.3% of the variance is measured in millions. For the median facility, one percent translates to more than $1 million dollars. For all facilities combined in 2003 dollars, one percent equates to $40 million.

It is interesting and of some concern to note the presence of an outlier, which would have affected all analyses (SFA and loglinear models with DEA scores). A plot of the leverage vector (which is nothing more than the trace of the "Hat" matrix, $X(X^TX)^{-1}X^T$)[9] versus Cook's D (a statistic for determining estimated *Y* outliers, see Cook and Weisberg, 1999)[10] reveals that one facility is highly influential: Facility T. Facility T, a MEDCEN, was also rated as Pareto-Koopmans efficient in all analysis. Since separate analyses are conducted for hospitals without MEDCENs, this finding is less disconcerting; nevertheless, future work might omit Facility T from analysis. (The facility is labeled as "Walter" in the following graph.)

[9] The "leverage" (the trace of the hat matrix) is a measure of regressor variance from the "center of mass" of the regressor geometry (the $(X^TX)^{-1})$ matrix).
[10] Cook's D is a measure of the Euclidean distance of the model estimate at point *i* versus the model estimate considering all the data. In simpler terms, the statistic is a distance measurement, which determines which estimates are *Y* outliers.

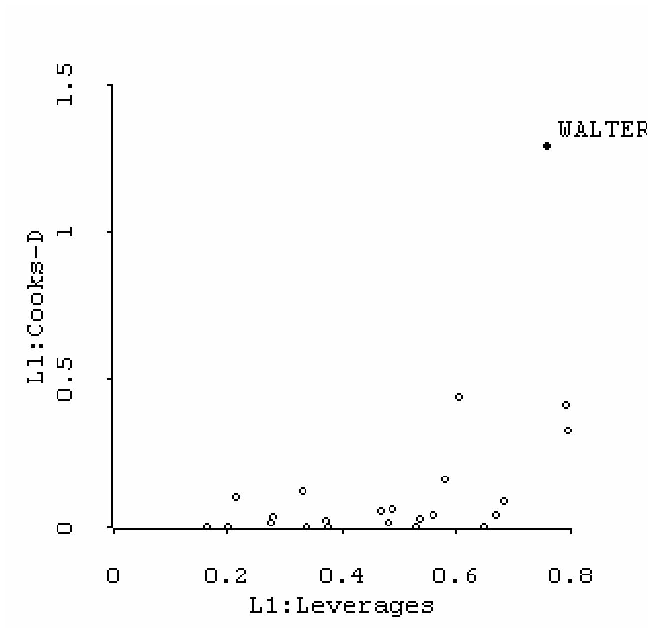


Figure 27. Leverages versus Cook's D, 2001 Regression.

The plot of leverages (the trace of the "Hat" matrix) versus Cook's D (a distance measurement for determining *Y* outliers) reveals an *X* and *Y* outlier in Facility T, which is labeled "Walter." The values for this DMU are highly influential in the fit the linear model. In fact, this analysis suggests that Facility T might be considered to be in a separate category as a super-MEDCEN without peers.

Linear model L2, which is now discussed, relies on the technical efficiency score.

Specifically, the model is formulated as specified in (66) from Chapter 4 but is provided below for convenience of the reader.

$$\text{Linear 2}$$

$$C = [1\ W\ P\ Q\ A\ E\ M\ M*W][\beta] + \varepsilon$$

$$E = \text{Technical Efficiency}$$

$$M = \begin{cases} 0 \text{ if not MEDCEN} \\ 1 \text{ if MEDCEN} \end{cases}$$

In this formulation, *E* is not treated as a dummy variable as was done in Linear Model 1 (L1). In this case, *E* is the technical efficiency score. Estimation is again performed through least squares applied to the n=24 observations in 2001. The results are reported in Table 46. Once again, one must interpret the results with caution due to the small sample size in comparison with the number of parameters. The results of the regression suggest a statistically significant model exists ($F(7,16)=142.128,\ p<.001$) with excellent R^2 (.984 unadjusted, .977 adjusted). Further, the workload and efficiency variables once again emerge as statistically significant. The directionality of the coefficients was proper (workload was associated with increased costs while efficiency was associated with decreased costs), and the effect size as measured by the standardized coefficients indicated that both variables contributed largely to the model ($W=1.007$, $E=-.213$). Additionally, the Quality variable was statistically significant in this model at the .05 level. The coefficient estimate suggests that increases in quality increase costs, which is exactly what one would expect.

Table 46. Linear Model 2 Regression Results, 2001.

The results of the regression are nearly equally impressive. The variance capture has improved to nearly 98% with fewer variables included in the model. The estimate for the variance is superior to the previous linear model (.217 versus .151). This model (with fewer variables) appears to be an improvement.

Summary measures

Multiple R	0.992
R-Square	0.984
Adj R-Square	0.977
StErr of Est	0.151

ANOVA Table

Source	df	SS	MS	F	p-value
Explained	7	22.6360	3.2337	142.1280	<.001
Unexplained	16	0.3640	0.0228		

Regression coefficients

	Coefficient	Std Err	t-value	p-value	Lower limit (5%)	Upper limit (95%)
Constant	0.0000	0.0308	0.0000	1.0000	-0.0653	0.0653
M	-0.0494	0.1011	-0.4888	0.6316	-0.2638	0.1650
MW	0.1066	0.1200	0.8885	0.3874	-0.1477	0.3609
P	-0.0788	0.0781	-1.0079	0.3285	-0.2444	0.0869
W* **	1.0072	0.1182	8.5241	0.0000	0.7567	1.2577
Q*	0.1030	0.0382	2.6978	0.0158	0.0221	0.1839
A	0.0551	0.0368	1.4992	0.1533	-0.0228	0.1330
E* **	-0.2126	0.0376	-5.6560	0.0000	-0.2922	-0.1329

******p<.001, *p<.05*

It is clear from the models above that Model L1 is preferred over Model L2 from a variance capture and standard error perspective, but, again, the relatively small sample size needs to be taken into account. Because the standard error of the estimate is smaller and because the variance capture is better, L2 is selected for additional analysis. It is useful to look at the model fitting plot for L2 (Figure 28). The fit is clearly linear and "tight," hence, L2 may be regarded as the better of the two models and select it for use in the ridge and robust regression models.

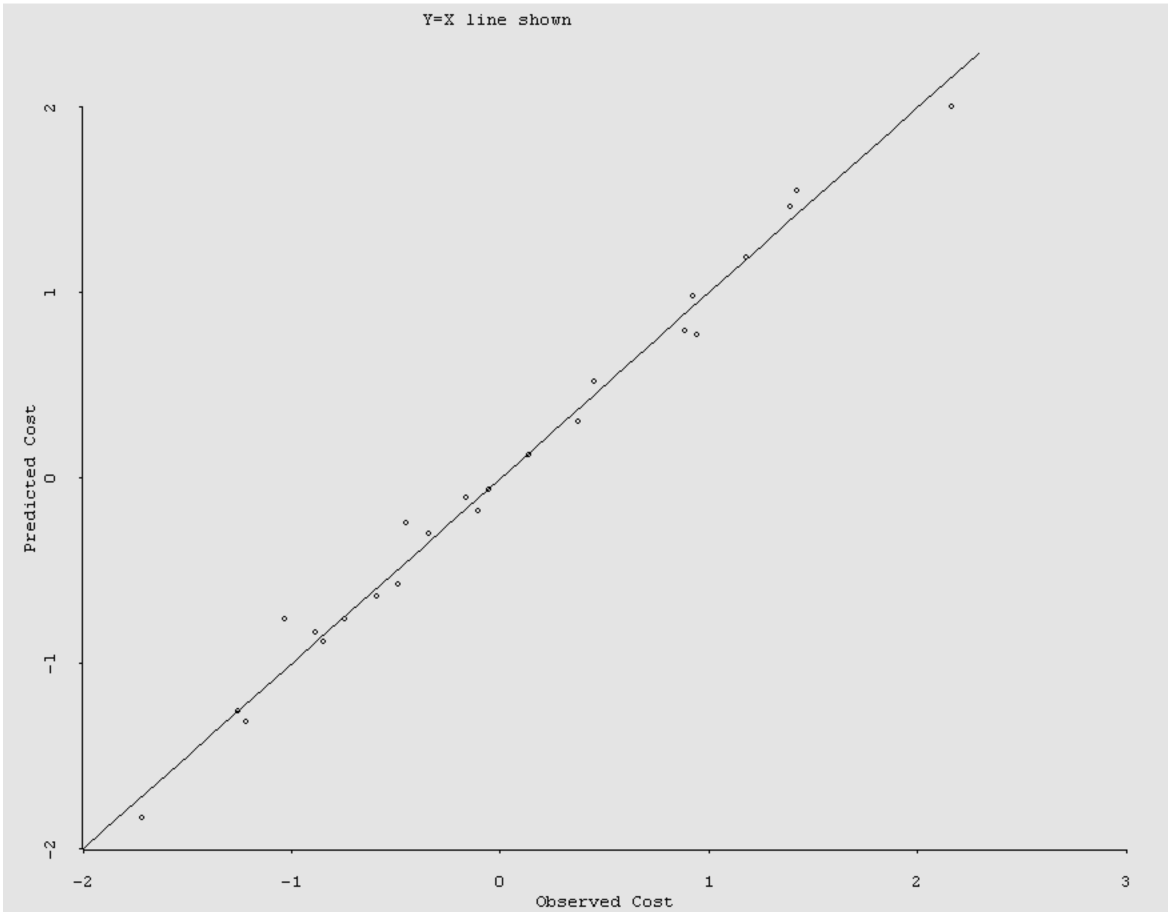


Figure 28. L2 Model Fit, 2001.

Bayesian Models

There are certain limitations associated with standard linear regression. Collinearity of regressor terms (which appeared to be nominal based upon the Variance Inflation Factors) still may have had an impact on the forecasts. For this problem, one might consider ridge regression. Also, the linear formulation did not allow for the fact that the predictors themselves might be drawn from a larger population of potential predictors.

For example, it factors like population and workload fluctuate based on uncontrollable events such as base realignment, increases in force structure, unit rotations into training posts, etc., and can affect the regressions. Robust regression should be considered. This logic leads to the formulation of three simple Bayesian models.

The initial Bayesian model (B1) is simply a reconstruction of the linear regression model using common priors on the distributions and assuming independent coefficients. The second Bayesian model (B2) assumes that the predictors co-vary, and therefore a common yet vague distribution is assigned to each coefficient's variance as in ridge regression (see Lindsley and Smith, 1972. In the third model (B3), it is assumed that the regressor terms were drawn from a larger population of potential predictors (robust regression).

All three models use the same form as Model L2 in the previous section (equation (66)), referred to here as the based model. The base model, assumptions, initializations, and terms follow. (Note: all variables are standardized prior to running the analyses.)

$$C = [1 \ W \ P \ Q \ A \ E \ M \ M * W][\beta] + \varepsilon$$

The assumptions of Bayesian Model B1 are $\beta \sim N(0, 10^4)$. In this case, each separate coefficient is modeled as an independent normal distribution with mean of zero and high but constant variance. In fact, the assumptions of independence and homoskedasticity are identical to that of traditional linear regression. The expected value and variance are updated during repeated simulation.

The assumptions of Bayesian Model B2 follow: $\beta \sim N(0, \eta)$, $\eta \sim IG(10^4, 10^4)$, *vague prior on* η. In this case, the assumption of homoskedasticity is dropped, and it is assumed that there exist different variances for each of the parameter vectors β. The variance of each of the parameters is assumed to be distributed as an inverted gamma with large parameters.[11]

The assumptions of Bayesian Model B3 follow: $\beta \sim MVN(\mu, \Sigma)$ *(Random Coefficients),* $\mu \sim MVN(\Delta, \Gamma)$, $\Sigma \sim$ *Wishart*$(\Lambda$, *# of regressors).* In this case, it is assumed that the parameter vector β has a multivariate normal distribution with mean vector μ and covariance matrix Σ. Further, distributions on both the mean vector μ and covariance matrix are taken. The traditional assumption is that the mean vector is also distributed as a multivariate normal random vector, and the vectors is provided vague initializations (not random vectors) associated with the terms Δ and Γ. Similarly, the covariance matrix takes on the usual multivariate Wishart distribution.[12] This formulation is equivalent to the assumption that the β parameters are random and is nothing more than a hierarchical model (see Lindsley and Smith, 1972). The initialization for Bayesian Model 3 (B3) follows: Δ *initialized as zero vector,* Γ *initialized as vague (precision of .01),* Σ *initialized with .1 correlation for all variables.*

[11] It is simple to illustrate that the distribution of the variance, σ^2, is Inverted Gamma based on the appropriate normality assumptions by transforming the well known distributional identity, $(n-1)S^2/\sigma^2 \sim \chi^2_{n-1}$.
[12] For simplicity, it is convenient to view the Wishart distribution as a multivariate version of the Inverted Gamma.

The terms remain identical to those in the linear models. That is the cost function is estimated for the traditional components of population, workload, quality, access, and efficiency.

An initial burn-in of 100,000 iterations was run for all three models in the WINBUGS program. WINBUGS (Windows Bayesian Inference Using Gibbs Sampling) is software designed by Medical Research Council Biostatistics Unit of Cambridge in conjunction with the Imperial College School of Medicine at St Mary's and is designed for Bayesian modeling. (See http://www.mrc-bsu.cam.ac.uk/bugs/welcome.shtml for more information). The code used is provided in Appendix 6. Autocorrelations were examined in all cases and were highly suggestive that convergence was achieved, as iteration dependence was nominal. Another 100,000 iterations were then performed to determine coefficient estimates. Estimates for the response variable were compared against actual values in order to evaluate model focus and to compare variance and bias. Further, the estimates for all models were compared against those attained through normal linear regression minimization of the squared error to validate the formulation. The coefficient estimates are nearly the same for all models, and the standard errors are very similar as well. A comparison of all three models is provided in Table 47.

Table 47. Bayesian Model Results, 2001.

A comparison of the parameter estimates for all models indicates the consistency. For all models, workload and efficiency are significant (*p<.01* for both). Note: The DIC term is the Deviance Information Criterion as specified in Appendix 8, MCMC Simulation. This term is useful in comparing the Bayesian models (lower is better). In this case, the assumptions associated with robust regression model would be preferred over the other ridge regression or traditional regression assumptions. Nevertheless, the coefficients are largely stable for all three models, and the directionality is identical.

	L2	Std Err	B1	SE	B2	SE	B3	SE
M	(0.049)	0.101	(0.019)	0.101	(0.010)	0.095	(0.019)	0.099
MW	0.107	0.120	0.079	0.120	0.122	0.112	0.093	0.117
P	(0.079)	0.078	(0.108)	0.082	(0.048)	0.086	(0.090)	0.081
W**	1.007	0.118	1.046	0.124	0.942	0.132	1.017	0.122
Q	0.103	0.038	0.076	0.038	0.080	0.039	0.078	0.039
A	0.055	0.037	0.038	0.037	0.034	0.037	0.038	0.037
E**	(0.213)	0.038	(0.208)	0.038	(0.198)	0.039	(0.205)	0.038
DIC			14.268		13.759		9.763	

***p<.01 for all models*

The Deviance Information Criterion (DIC) is smaller for B3 (as shown in Table 47), indicating that the robust regression model is preferred; however, a comparison of the observed variance and bias of each of these estimators suggests that all are very good. Table 48 provides a comparison of the mean bias and variance for models L2, B1, B2, and B3. While B3 had the smallest DIC, model B1 exhibited the smallest empirical variance and mean bias for the 2001 data.

Table 48. Comparison of Bias and Variance, 2001 Bayesian Models.

The table below provides a side-by-side comparison of the empirical variance and bias attained from the models after 100,000 iterations of MCMC. It appears that the simple linear regression model is a sufficiently robust tool for this dataset.

	L2	B1	B2	B3
Mean Bias	0.00000	0.00001	(0.00002)	0.00002
Variance	0.01517	0.01507	0.01581	0.01516

Formal pairwise tests of the Bayesian hypotheses are not necessary in this case, as all three models provide nearly the same estimates and the simplest formulation appears to be the best. With this analysis complete, it appears that the linear transformations and least squares solutions may provide reasonable forecasting models.

Model Comparison

A comparison of the linear models L2 and B3 with the SFA models S1 and S2 suggests that the SFA models are inferior based on the variance and bias as illustrated by Figure 29. The graph depicts the residuals on the *Y-Axis* against the DMUs on the *X-Axis*. (A perfect model would have all estimates for the DMUs located on the *X-Axis*.) It is clear from the diagram that the residuals for both SFA models are nearly always positive and that the variance for these models is significantly larger than that of the loglinear models with DEA efficiency scores solved either through traditional or robust regression. Because of the compelling visual evidence provided by Figure 29, formal hypothesis tests of the model differences are not necessary. (Note: for thoroughness, the SFA-HN model is retained for use in the next section in order to evaluate its forecasting capacity.)

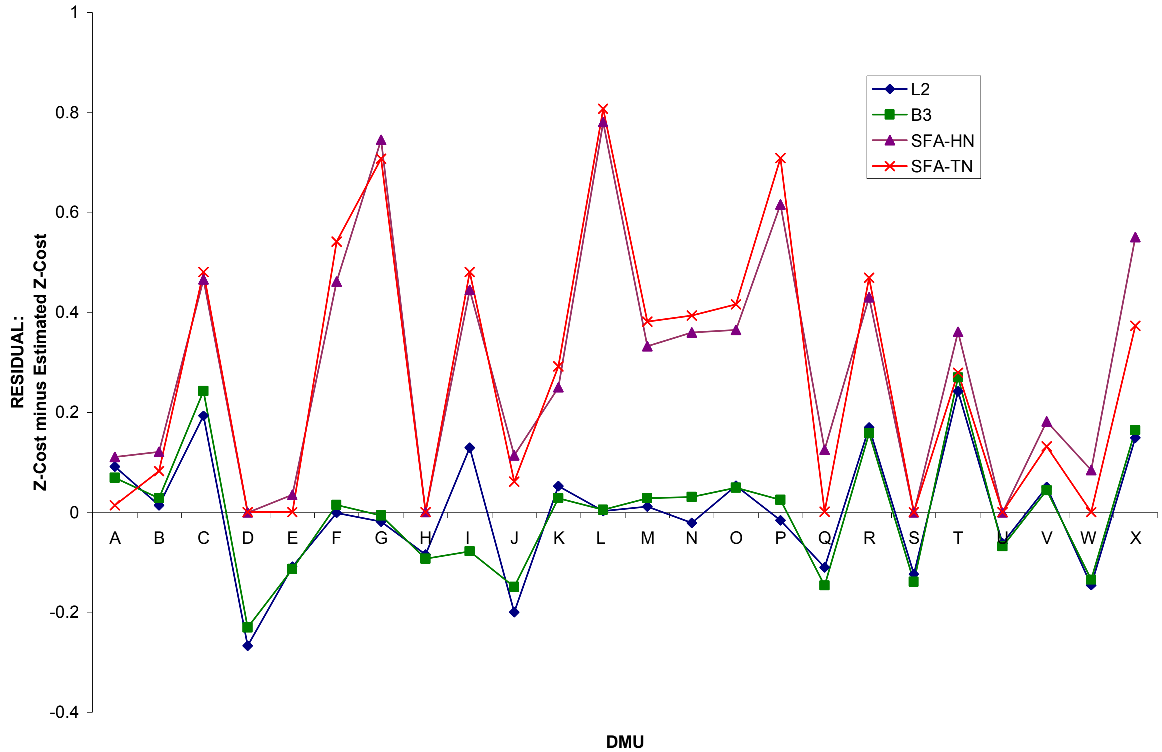


Figure 29. Comparison of Residuals: L2, B3, and SFA Models, 2001.
The linear models clearly perform better. B2 is omitted as its points are unobservable (obscured by L2).

<u>Forecasts for 2002</u>

With the initial analysis of the models in hand, it is simple to evaluate the forecasting efficacy to see if useable results might be obtained. In practice, decision-makers use the current year performance to plan for future budgeting. Empirically, one applies the estimates obtained from the previous analyses to the 2002 dataset.

The methodology for this section is simple. Variable transformations are performed on the 2002 dataset using the coefficients generated by the 2001 observations. Specifically, all variables from the 2002 dataset are transformed identically as in shown in Chapter 4 equations (69), (70), (71), and Table 13 with the coefficient estimates generated from the 2001 dataset. In this way, only the information available to decision-makers in 2001 is employed for forecasting data in 2002.

After the variables are transformed, an assumption regarding efficiency must be made. With the cross-sectional analysis, it is reasonable to assume that inefficiency should remain nearly the same over a short span of time. This assumption is relaxed later for panel analysis, where a moving average approach is adopted. With this assumption, estimates of costs are obtained and compared to the observed costs. The models with the smaller deviations from the observations are preferred. These models are then used as resource allocation decision support tool.

It is important to underscore the fact that the SFA efficiency estimates for 2001 are included in the forecasting model for 2002. Specifically, the inefficiency scores are converted back into the random variable u by taking the natural logarithm (since $CE = \exp\{-u\}$) in the loglinear SFA models.

The SFA models were nearly identical, so the half-normal model was the only one used to analyze its predictive capacity. The L2 loglinear model produces results nearly identical to L1, so it is selected for use as well. The estimates produced by robust regression (B3) were also used, as the B3 model was associated with the smallest Deviance Information Criteria.

After transforming the 2002 variables using the 2001 data (see equations 69-71 and Table 13 for exact specifications of the transformations), estimates for both the SFA and the linear models were derived. These estimates are depicted graphically in Figure 30 by way of a residual graph constructed the same way as in Figure 29. This graph provides clear visual evidence that indicates L2 results in the smallest residuals and that these residuals are clearly centered at zero. The implication is that the best forecasting model is L2. The SFA-HN model is clearly inferior to either L2 or B3.

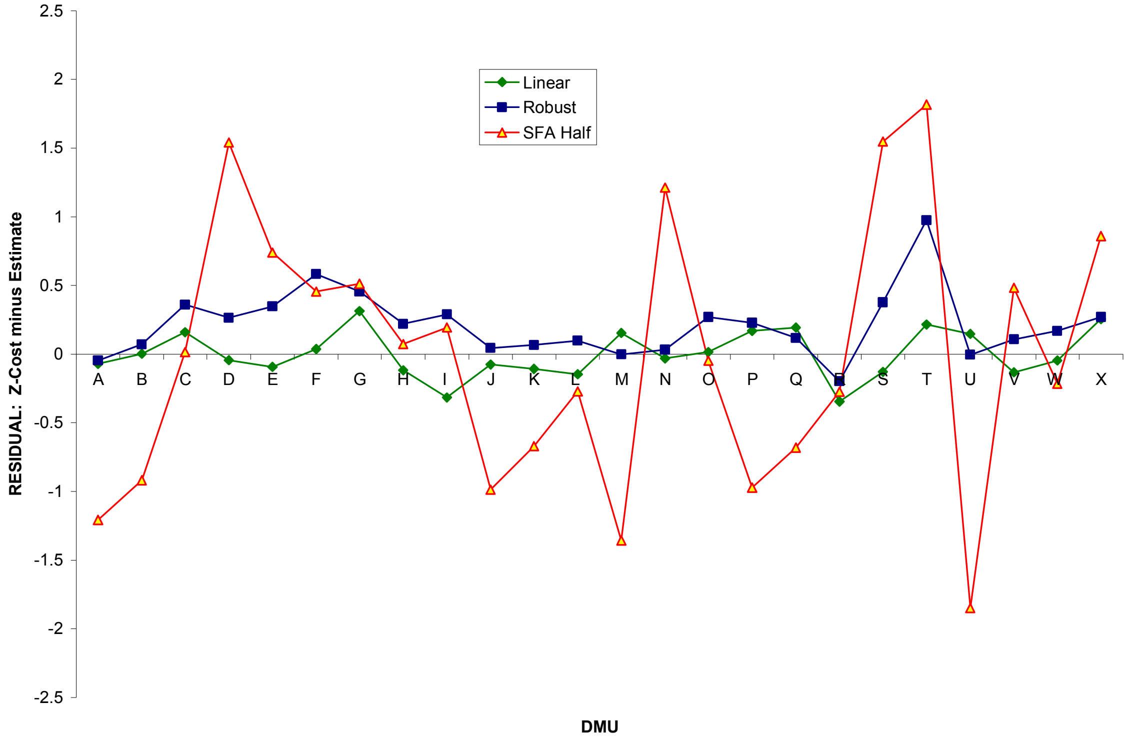


Figure 30. Comparison of Forecast Residuals: 2001 onto 2002.
From observation, the linear model appears to provide the best forecasts.

A comparison of the observed mean bias and variance for each of the estimators is provided in Table 49. From the table, one can readily see that the smallest bias and variance are associated with L2.

Table 49. Forecasting Bias and Variance, 2001 on 2002.
The chart below shows that L2 is the best estimator based on mean bias and variance.

	L2	B3	SFA-HN
Mean Bias	0.0000	0.2123	0.0000
Variance	0.0280	0.0552	0.9238

The linear model loses two percent variance capture in forecasting; however, the efficiency term retains its significance and variance share. When the coefficient associated with efficiency is set to zero in the forecasting model, the variance for the forecast drops nearly two percent. Without any further analysis, it is clear that, for the cross-sectional data, the linear model provided the best forecasts of the models evaluated. Using the forecasting models for management adjustment of resources will be discussed later.

Panel Data Analysis, 2001-2002

The cross-sectional analysis and forecast provided some insight into appropriate models, but a use of panel data models may improve the forecasts. In this section, an analysis of efficiency for both 2001 and 2002 is performed using DEA and SFA. The SFA model is evaluated as a time varying model as proposed by Lee and Schmidt (1993), while the linear models associated with DEA are solved via least squares. The robust regression problem (B3) is solved by MCMC.

As before, variables were transformed as if the 2003 data did not exist. The factor analysis for the *Population* factor for 2001-2002 data produced the following equations:

$$Factor_P = -1.719 + 0.0216*Enrollment + 0.0120*Population.$$

Again, this factor was shifted right 1.5 units and the logarithm was taken. Similarly the *Workload* factor for the 2001-2002 data became:

$$Factor_W = -.300 + 0.014*RWPs + 0.002*RVUs$$

These equations were then used to produce the regressor transformations for the dataset.

<u>Model Comparison</u>

Using the same techniques employed previously, it is simple to determine that the linear model provides nearly the same estimates as the Bayesian B3 model. Figure 31 illustrates this point and also shows how poorly the SFA-HN model performs when used in a forecasting model with assumption of time invariant efficiency. The graph shows the residuals on the *Y-Axis* and the DMUs on the *X-Axis* as before.

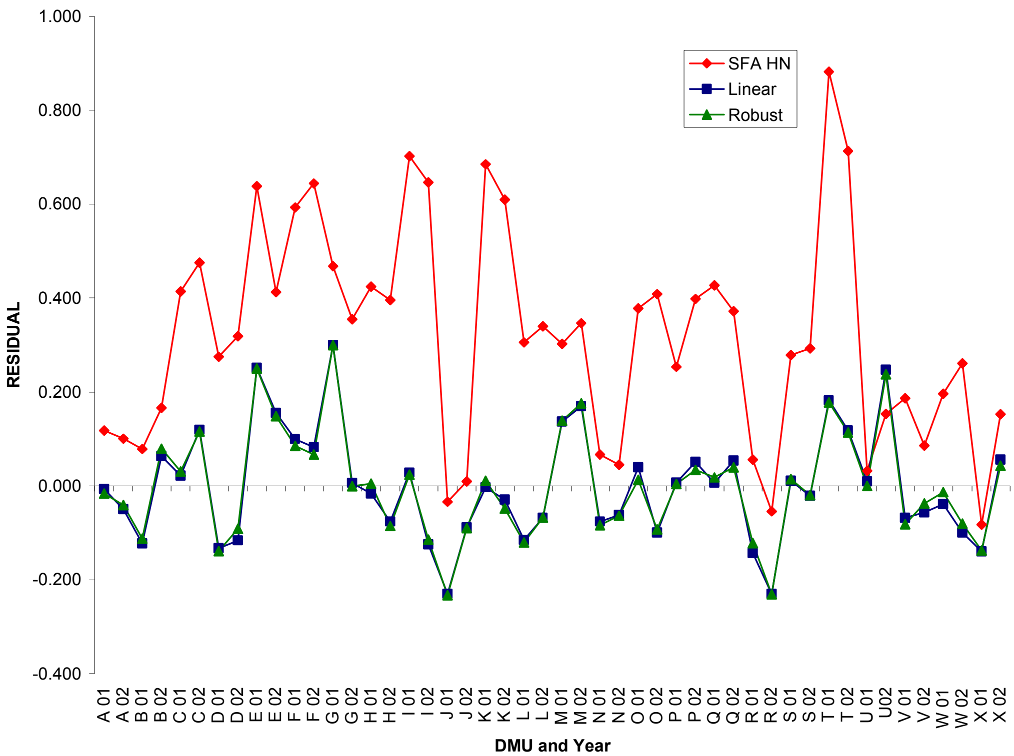


Figure 31. Comparison of Residuals L2, B3, and SFA, 2001-2002.
The graph above depicts the observed cost function, the SFA-HN estimate, the L1 estimate, and the robust regression estimate. Of the models investigated, the robust regression model provides a slightly better solution than the linear model solved through least squares; however, this is not readily apparent by graphing.

Based on the graphical illustration of the estimators, the comparison will be reduced to the L2 and B3 models. The regression analysis of the L2 model shows that is statistically significant (*$F(8,39)=341.5297$, $p<.001$*) with R^2= .983. Table 50 depicts the regression for L2. One immediately notes that workload and efficiency remain statistically significant (*p<.001* for both) and that the interaction term for MEDCEN*Workload and the quality proxy enter the equation as well (*p<.05, p<.001* respectively). The effect size of workload and efficiency are nearly identical to that determined in the Year 2001 analysis and the directionality is appropriate. Based on the results of the regression, an increase in the quality / prevention proxy results in an associated increase in cost. This finding is consistent with economic theory. Further, the interaction effect is directionally proper with expectations. Specifically, increasing workload in MEDCENS increases costs more dramatically than increasing workload in ACHs.

Table 50. Regression for L2.

The results are statistically significant ($F_{(8, 39)} = 341.533$, $p<.001$, $R^2=.98$).

Summary measures

Multiple R	0.9929
R-Square	0.9859
Adj R-Square	0.9830
StErr of Est	0.1302

ANOVA Table

Source	df	SS	MS	F	p-value
Explained	8	46.3386	5.7923	341.5297	<.001
Unexplained	39	0.6614	0.0170		

Regression coefficients

	Coefficient	Std Err	t-value	p-value	Lower limit (5%)	Upper limit (95%)
Constant	0.0000	0.0188	0.0000	1.0000	-0.0380	0.0380
M	-0.0371	0.0598	-0.6209	0.5383	-0.1580	0.0838
MW*	0.1700	0.0731	2.3242	**0.0254**	0.0221	0.3179
P	-0.0780	0.0471	-1.6573	0.1055	-0.1733	0.0172
W***	0.9612	0.0699	13.7510	**0.0000**	0.8198	1.1026
Q***	0.0873	0.0209	4.1672	**0.0002**	0.0449	0.1296
A	0.0399	0.0218	1.8353	0.0741	-0.0041	0.0839
E***	-0.2088	0.0244	-8.5568	**0.0000**	-0.2582	-0.1595
T	0.0205	0.0195	1.0507	0.2999	-0.0190	0.0600

*$p<.05$, ***$p<.001$

Similar detail for the B3 model is not provided because the coefficient estimates are nearly identical as seen in Table 51. The differences exist primarily at the second and third significant digits.

Table 51. Coefficient Comparison, 2001-2002.

The following table illustrates that L2 and B3 models are nearly identical.

	L2		B3	
	Coefficient	**SE**	**Mean Est.**	**SE**
M	-0.037	0.060	-0.043	0.059
MW* **	0.170	0.073	0.176	0.072
P	-0.078	0.047	-0.080	0.047
W* **	0.961	0.070	0.962	0.069
Q* **	0.087	0.021	0.085	0.021
A	0.040	0.022	0.041	0.021
E* **	-0.209	0.024	-0.211	0.024
T	0.021	0.020	0.017	0.019

A table of the empirical mean bias and variance (Table 52) illustrates that the robust regression estimates are slightly better than the L2 estimates and both are better than the SFA results which are biased and have higher variance.

Table 52. Variance and Bias of Estimators, 2001-2002.

The robust regression results in the smallest variance and zero bias, although the LSDV model is very close.

	SFA-1 (HN)	L2 (LSDV)	B3 (Robust)
Mean Bias	0.319	0.000	0.000
Variance	0.153	0.014	0.013

Forecasting

Using a similar methodology as the panel data employed, forecasting of the 2003 period was performed. Instead of assuming time invariant efficiency, a weighted moving average approach was used (Greene, 2003). This approach assumes some time-related efficiency effects. Forecasted efficiency scores were then generated as follows:

$E_{2003} = \frac{2 * E_{2002} + E_{2001}}{3}$. In this formulation, E_{200i} is the efficiency score for the year 200i, so that the end result is a simple weighted moving average. The efficiency scores generated in the year 2002 are considered to be twice as relevant as those generated in the year 2001 in determination of the forecast for 2003.

The results of the forecast are provided in Figure 32. From observation it is apparent that L2 and B3 are nearly identical. Further, it is interesting to note that efficiency scores should be retained for the forecasting models as well, since they are associated with more than two percent of the variance capture in the forecast.

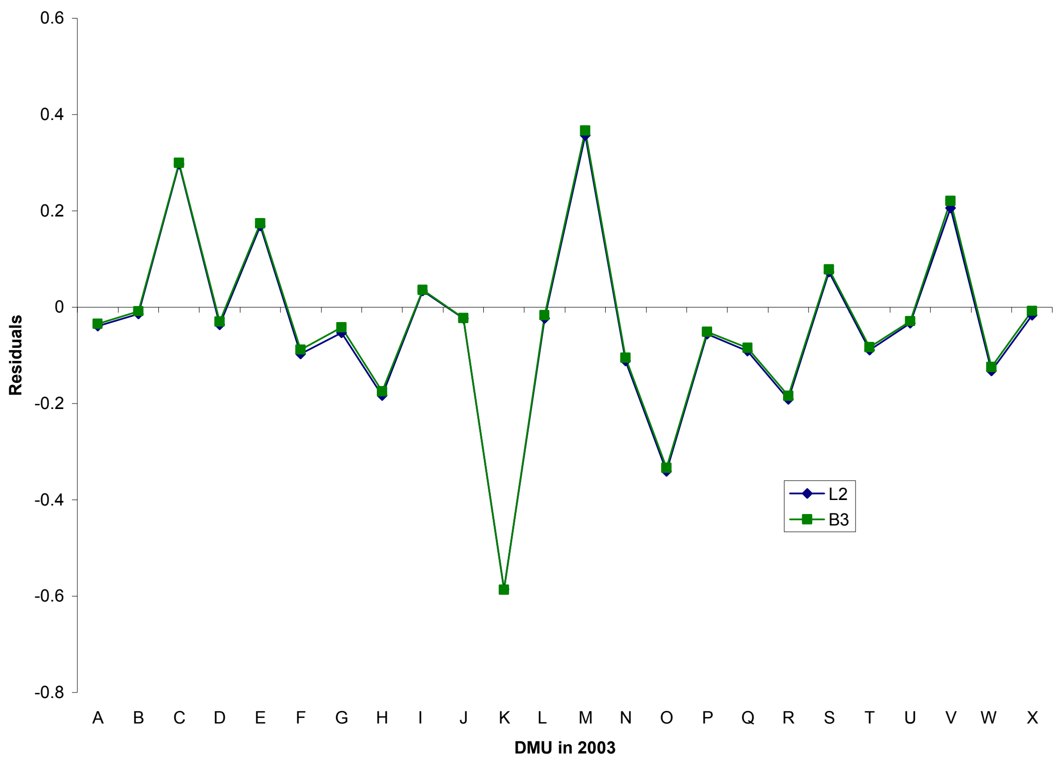


Figure 32. Comparison of Forecasting Residuals, 2001-2002 onto 2003.
The figure above illustrates that L2 and B3 provide nearly identical estimates.

Resource Allocation

One of the major purposes of this dissertation was to develop performance analysis for AMEDD components that might be used for resource allocation. Models have been tested that seem to have reasonable forecasting capability for cost based upon assumptions of efficiency. Use of these models requires management input.

One method for using these models follows.

1. Efficiency scores for time *t* are evaluated exactly as above (either cross-sectionally or using panel data, if available).
2. Using the estimates of efficiency at time *t*, an initial forecast of costs is generated. This initial estimate is a simple forecast as performed above without adjustments for inefficiency, but it includes a manager set error level (e.g., one standard error.)
3. Next, a revised estimate is generated. Facilities operating below a manager set efficiency score benchmark (e.g., 98%) are adjusted upwards. For example, if the facility achieved 95% efficiency over the past two years based on a moving average approach, the score would be shifted upwards 3% when calculating a revised forecast.
4. Revised budget estimates are then projected to time period *t*+1 as if the facility were operating at the benchmark.
5. If the revised estimates fall outside the manager set error level associated with the initial estimates, then reallocation should be considered.

An illustration of this approach is provided in Figure 33, which forecasts 2003 estimates and revised estimates assuming 98% efficiency. The graph provides lines for the 2002 budget (standardized), the 2003 budget (standardized), and the revised estimate based on a 98% management directed efficiency minimum. The revised estimate would have recommended decrements for the following facilities based solely on the 2002 data: M, K, I, C, V, and E. These facilities lie below the 2002 values and are also below the managerial set error (in this case, a 10% standard error band). Further, increments would have been recommended for Facilities R, Q, W, H, O, and G. It is interesting to note that the actual funding of the facilities (as indicated by the 2003 "actually funded" line in Figure 33) resulted in major decrements for Facilities K and F, while major increments were provided for A, J, R, W, I, and C. The only similarities between recommended funding and actual funding would have been Facilities K, R, and W.[13]

[13] In every efficiency analysis and cost model, Facilities I and C were inefficient. Despite the inefficiency, the facilities received moderate increments in 2003.

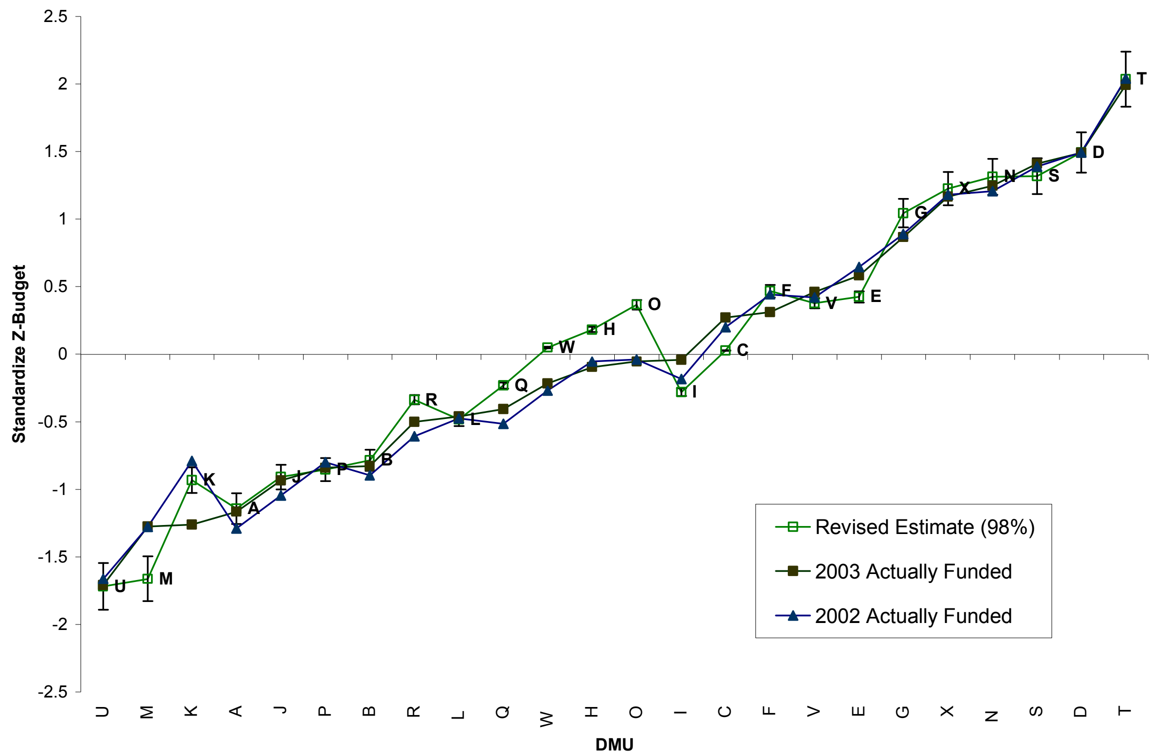


Figure 33. Resource Allocation Using L2, 2001-2002 onto 2003.
The graph above provides the actual costs in 2003, the estimated cost based on a weighted average forecast, and the estimated cost based on a management directed efficiency floor.

The recommendations associated with the analysis of this graph are similar to those provided by a separate DEA cross-sectional analysis of the year 2003. One notices that three facilities that might be decremented (I, V, and C) are also have the largest cost projections in the BCC-I analysis of year 2003 as shown in Table 53. This result is interesting and serves to provide support for the forecasting methodology.

Table 53. BCC-I Analysis, 2003.

The cost projections for the 2003 cross-sectional analysis suggest that the facilities identified for decrement in the forecasting models may indeed be over-funded. Specifically, facilities I, V, and C were identified for decrements. The plus-ups are not obvious from the traditional BCC-I formulation; however, it is possible to determine superefficient units by reformulating the mathematical program without the inclusion of the evaluated DMU in the constraints. This method will be evaluated in future research.

Facility	Cost Projection
I	-10.95%
V	-7.09%
C	-6.49%
P	-6.31%
G	-3.98%
B	-3.11%
W	-2.90%
L	-2.28%
X	-2.21%
K	-2.07%
F	-1.95%
H	-1.07%

The potential plus-ups are not fully congruent with the DEA analysis, however. Specifically, the forecasting model identifies R, Q, W, H, O and G for potential increases in budgeting. Of these facilities, the DEA BCC-I analysis identifies three as fully efficient (Q, R, and O).

Summary and Conclusions for Cost Models

In this chapter, multiple cost models for individual and grouped years were investigated. The loglinear models with efficiency scores (regardless of type) performed universally better than any Stochastic Frontier Analysis cost model. In fact, the best linear models evaluated were statistically significant ($p<.001$ for all models) estimated with Ordinary Least Squares and resulted in $R^2=.98$. Most impressively, the models were capable of forecasting from time t to time $t+1$ with excellent accuracy ($p<.001$, $R^2=.95$ for both cross-sectional and panel series forecasts). Finally, adjusting the efficiency scores to a managerial set minimum and using these scores to recommend resource allocations in year *t+1* proved to be provide results consistent with the recommendations of a separate DEA analysis of the year *t+1*, as Facilities I, V, and C are seen as highly inefficient in the DEA analysis and are also seen as requiring decrements in the cost models. The results of the cost models provide some support for use of the DEA BCC-I slacks for analyzing facility budgets

From the analysis, there is sufficient evidence that workload and efficiency factors are the largest drivers. In fact, 94% of the response variable appears to be related to workload, while 3% is related to efficiency in the best linear model. Quality and other factors compose the remaining small portion of important regressors. The cost models indicate that the MEDCEN interaction with workload is an important regressor term as well.

With the efficiency and cost analyses complete, one may turn attention to summarizing the results and concluding this research. These two objectives are the focus of Chapter 8.

Chapter 7: Summary, Limitations, Recommendations, and Conclusions

The objectives of this dissertation were to examine possible approaches for evaluating cost, quality and prevention, access, training, readiness, and workload performance for AMEDD components in order to make recommendations for possible resource shifts within and across military medical facilities that have a total annual budget of approximately $4 billion. The multi-method process is now summarized.

Method Summary

The literature revealed multiple performance-based resource allocation methods which have been applied in the public sector (and in healthcare settings specifically). The selection of a single methodology, however, would possibly result in methodological bias (Charnes, Cooper, and Sueyoshi, 1988), and so multiple methods were employed.

Previous models seemed to suggest that it would be possible to select a parsimonious set of variables, which would account for complexity, volume, and efficiency measures (Foster and Gupta, 1990). Research such as Charnes et al (1985) and Coppola (2003) provided the foundations for developing inputs and outputs, which were used in both efficiency analyses and in loglinear models. The primary variables selected for inclusion in the study were facility total budget / cost, Relative Weighted Product, Relative Value Units, a quality / prevention proxy from the Health Care Survey of Defense Beneficiaries, and an access proxy from the same survey. Because these variables were to be used in parametric analysis, they were transformed using factor analysis and the natural logarithm to ensure that they met the appropriate assumptions. The primary model used for the parametric analysis suggested that facility budgets could be properly determined by measuring a single workload variable, an efficiency variable, a quality variable, the facility's medical center status, and the interaction between the workload variable and medical center status. Different assumptions regarding this model were made in order to determine the most appropriate assumptions for making forecasts.

The efficiency variable was measured using several different methodologies. Efficiency was estimated by Data Envelopment Analysis, Stochastic Frontier Analysis, and Corrected Ordinary Least Squares. These methods each provided different albeit correlated measures of efficiency. Careful evaluation suggested that either the DEA or SFA efficiency scores should be retained in the budget / cost models. .

The SFA procedure provided cost estimates; however, the DEA procedure did not. To generate these estimates, scores produced by the DEA efficiency analyses were then included in two separate linear regression models. The first model used Pareto-Koopmans efficiency scores as binary variables, while the second model used technical efficiency scores as continuous variables. The assumptions regarding linear regression models were then relaxed by using Bayesian hierarchical models (ridge and robust regression) and estimating the parameters via Gibbs sampling.

In order to evaluate the forecasting capacity of the models, the three-year dataset (2001-2003) was split into groups. First, analysis was performed on the 2001 data and applied to the 2002 data by means of forecasting. Efficiency was assumed to be time invariant in order to forecast costs. Second, analysis was performed on the 2001-2002 data and applied to the 2003 data. Efficiency was modeled as a simple weighted moving average. An analysis of all three years of data was also performed.

Finally, a method for determining resource allocation was offered. This method involved setting managerial specified efficiency ratings and error bands. A hospital's budget was then forecast as if it was at least as efficient as the managerial set rating (98%).

Results Summary

The results of the efficiency analysis were congruent in identifying primarily the low and high performers. Most notably, Facilities F and I were near the bottom in all analyses for all three years, while Facility K was in the bottom two out of three years. From the DEA analysis, multi-million dollar cost adjustments would have been recommended each year. The cross-sectional analyses for the separate methodologies were more similar than the multi-year analyses (e.g., 2001 SFA and DEA results were more highly correlated than 2001-2002 SFA and DEA results). An analysis of the low and high performing DMUs revealed no obvious relationships, which might have indicated omitted variables.

The analysis of the cost models revealed that loglinear models with DEA efficiency scores routinely exhibited less mean bias and variance and were better in forecasting. The best models, L2 and B3, were statistically significant ($p<.01$ for both models) and captured 98% of the variance when applied internally and 96% when used to forecast future costs. The efficiency variable was retained in all models. Based upon the forecasts produced by the loglinear model, decrements for facilities I, V, and C would have been considered for the year 2003 along with increments for facilities K, R,Q, W, H, and O. The results were similar to those found by the BCC-I (Input Oriented) analysis of the 2003 data.

Limitations

While the models presented herein show potential, the variable selection may require further study and improvement. The quality proxy measure is primarily related to preventive health (which is also a measure of quality), but it might be improved by the addition of other metrics. Further, the population metric does not reflect "reliant users," those persons that routinely use the healthcare facilities (three times per year or more). This metric is relatively new and might improve the contribution of population in the models. Instead, the current population regressor is a measure of local demography and ignores the rotation of trainees into bases and other factors. (Use of robust regression, which treats these regressor terms as random variables, provided a method for partially addressing this issue, however.)

Further, it is noted that a universal measure of efficiency must be approved by the command structure of the Army Medical Department. As illustrated plainly in this dissertation, efficiency scores differ depending on the model selected. The ability to obtain reliable estimates of efficiency is clearly a function of appropriate model selection.

Next, the selection of models was limited by the data available in the short panel series. If monthly or quarterly data for all variables becomes available, it would be possible to expand the panel series such that the feasible GLS estimation might be used. With short panels, dummy variables suffice and reasonable results were obtained.

Recommendations and Future Research

For the population of hospitals that was investigated, about $120 million annually may be attributed to inefficiency. While small in comparison to the combined budgets of these hospitals ($4 billion), the amount does represent a significant potential for improving system performance. Based upon these results, Facilities F and I deserve additional scrutiny.

Future research should focus on network to MTF comparisons. The foundation for this analysis is provided in Appendix 8, which provides the following findings.

1. The MTFs are universally less efficient than the networks based on the variables selected, which was to be expected due to the readiness and training requirements unique to military facilities.

2. Inefficient networks were generally associated with efficient MTFs and vice versa.
3. An analysis of 25 individual network services provided a starting point for research regarding the allocation of manpower to MTFs and referral patterns for the AMEDD. The results suggested that the mean efficiency of nutrition services and hematology were the lowest of all evaluated, and that Facility I had the lowest mean average efficiency score for the services evaluated.

Additional joint network and MTF analysis should be expanded to determine, which network components are efficient and capable of handling more demand. Inefficient network components should be targeted for work recovery in order to move high cost services into the military treatment facilities.

Further study of low and high performing DMUs is required to understand the proper components of efficiency models. Cost excesses, quality shortages, and access shortages may be related to missing variables (and in some cases, this may be the case.) In other cases, however, a detailed analysis of the low performing DMUs may substantiate the model findings. In these cases, resource recommendations are available from both the cost models and the DEA slacks.

The analysis of performance should be shared with AMEDD resource managers, improved, and automated so that the leadership has a reasonable ability to make budgeting decisions outside of the historical plus inflation approach. Models that link resources to performance should serve to balance the AMEDD's provision of care to its authorized beneficiaries.

Further research should analyze superefficiency scores as a possible method for determining required cost increments. Specifically, super-efficient DMUs may actually represent either underfunded or overworked DMUs.

Another line of research would disaggregate the costs and evaluate the efficiencies and cost models by type. While disaggregation is somewhat artificial in AMEDD facilities, which use stepdown accounting procedures that smooth costs by often uncontrollable measures[14], it might serve to provide additional insight.

In general, the use of targeted Operations Research studies should be expanded to handle the $8 billion AMEDD industry. The need for trained Operations Researchers is clear.

Conclusions

The aim of this dissertation was to investigate performance of AMEDD military facilities and to link that performance to resource allocation. Specifically, the goal of the dissertation was to evaluate cost, quality and prevention, access, training, readiness, and workload performance for AMEDD components in order to make recommendations for possible resource shifts within and across military medical facilities. By coupling non-parametric and parametric methods, performance measures were effectively linked with resource allocation models. From these models, forecasts were made and

[14] For example, maintenance costs are parceled out to each department, clinic, and service based on square footage, which is not normally a controllable factor.

validated. The models provided reliable estimates of costs based upon efficiency assumptions. The goals of the dissertation have been met.

The challenge, however, remains in the modification of the work so that AMEDD leaders accept and use it as a decision support tool for reallocation or facility efficiency measurement. By careful coordination with the AMEDD senior leadership, revision of variables, and continued model testing and validation, the analysis provided herein may prove useful in budgeting the $4 billion costs associated with the 24 facilities in this study. It may also be useful in improving the efficiency of the AMEDD and Network facilities, although much further work is needed to achieve this objective.

Appendix 1-Efficiency Software Comparison

The one phase GAMS formulation discussed in Chapter 3 assigns the non-Archimedean element a value of 10^{-4}. The GAMS formulation for the software comparison is in Appendix 2. The advantage of this formulation is that the slacks and the technical efficiency scores are obtainable in a single LP solution. While instability can occur in this one-phase formulation, it is interesting to note that this problem is not observed with this sample data.

The two-phase GAMS formulation is congruent with the recommendations of Cooper, Seiford, and Tone (2000) in avoiding the potentially unstable one-phase formulation. In the first phase, a technical efficiency score is generated by solving the simple linear program specified below, in which the objective does not involve the slacks.

$$
\begin{aligned}
&\text{Minimize } \beta \\
&Y\lambda \geq y_i \\
&X_D\lambda \leq \beta x_i \\
&X_{ND}\lambda = x_i \\
&e\lambda = 1 \\
&x \geq 0,\ y \geq 0,\ \lambda \geq 0, \beta \geq 0
\end{aligned}
$$

Let β^* be the optimal value of the β in the above LP. In phase 2, the slacks are maximized by solving the next optimization problem. In this formulation, the input slack is restricted to the discretionary inputs. The non-discretionary inputs do not have slack and are therefore excluded from the objective function.

$$
\begin{aligned}
&\text{Maximize } w = es^- + es^+ \\
&es^+ = Y\lambda - y_i \\
&s^- = \beta^* x_i - X_D\lambda \\
&e\lambda = 1 \\
&x \geq 0,\ y \geq 0,\ \lambda \geq 0, \beta \geq 0
\end{aligned}
$$

Appendix 3 illustrates the GAMS model for the two-phase model.

DEA Solver allows a user to formulate and solve several DEA models. The user formats the data on an Excel spreadsheet and then runs the Visual Basic for Applications macro, which is DEA solver. While the details of the algorithm are not available for investigation, the results match both GAMS formulations identically, lending credence to the GAMS models. A snapshot of one of DEA Solver's forms follow.

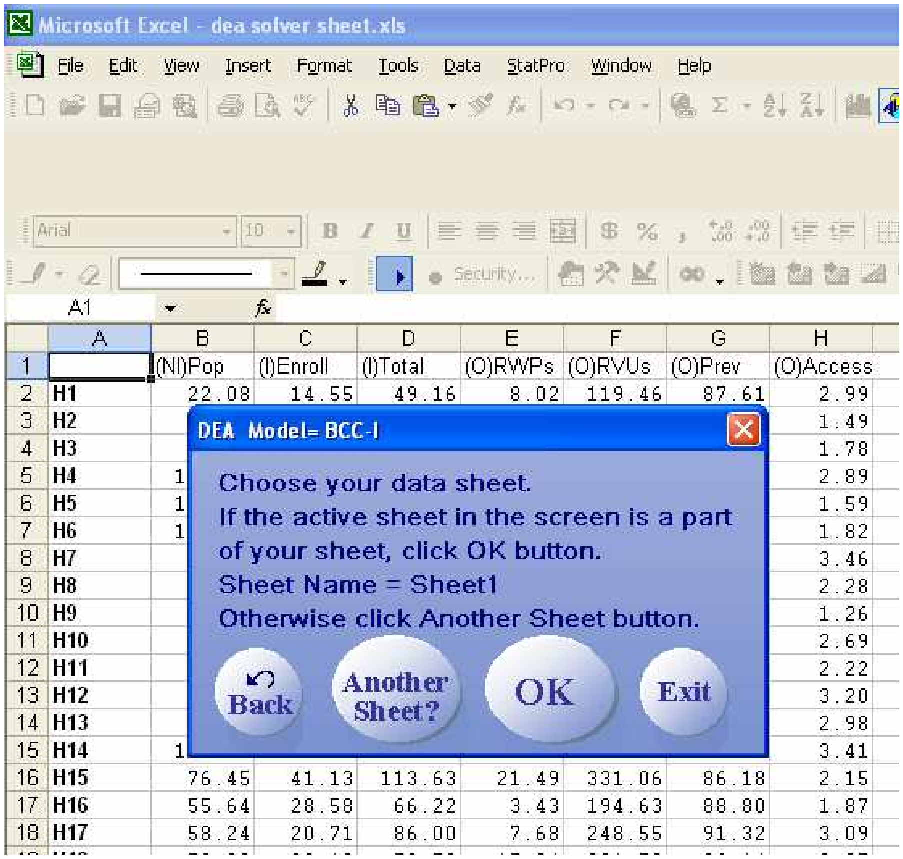


Figure 34. DEA Solver Screen Snapshot.

The DEA Solver interface is based on the popular Microsoft Excel and Visual Basic for Applications format. The spreadsheet depicted above reflects the non-discretionary variable "Population" annotated with the initials "(NI)." Other outputs and inputs are annotated accordingly.

The EMS software has data input format similar to DEA Solver, although the data must be prepared outside of the software. This software also allows the user to select from multiple models and formulations (including times series) but does not appear to be as robust as DEA solver, i.e. its output is not identical to that from both GAMS models and the DEA Solver software, for some problem instances. A screen snapshot follows.

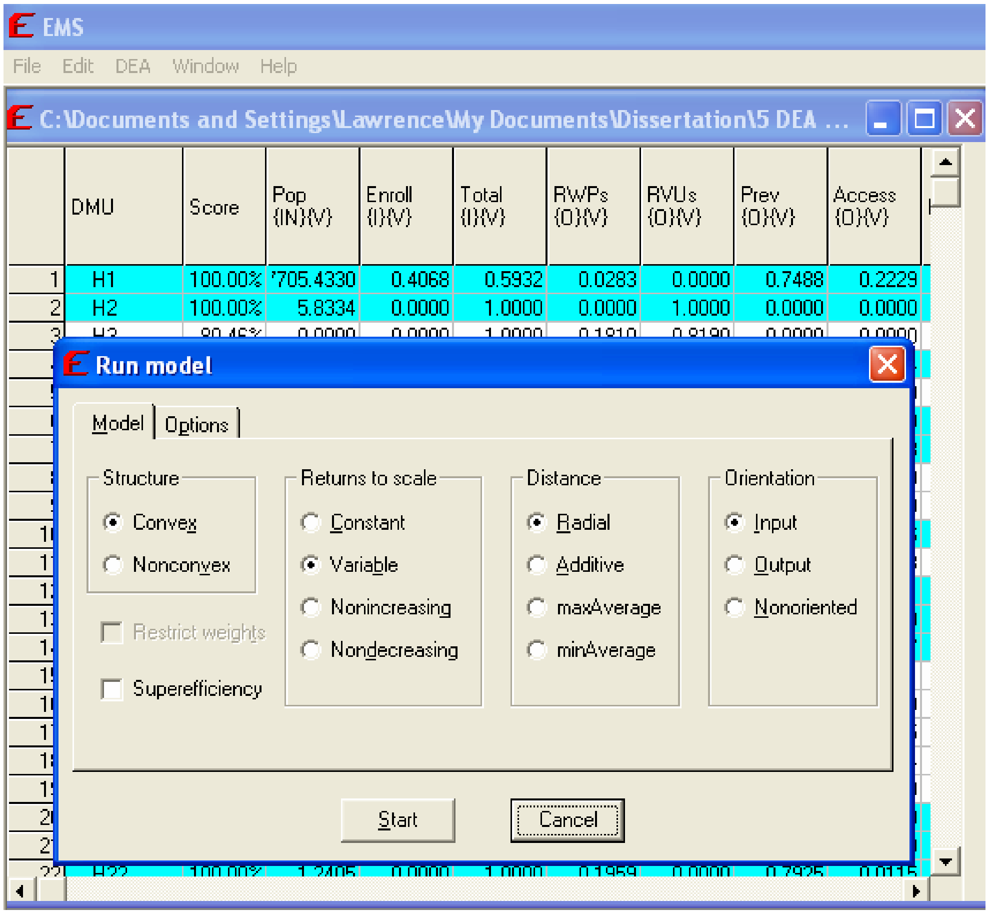


Figure 35. EMS Screen Snapshot.

The EMS input screen and associated results are depicted above. Data are formatted outside of the software. Similar to the DEA Solver software, there are multiple model choices. The spreadsheet produced is readily transportable to Excel. The output is not as robust as DEA Solver.

A comparison of the technical efficiency scores generated by the models for the trial analysis is provided in Appendix 3. One quickly notices that there are seven DMUs out of 72, in which the efficiency scores generated by the EMS variable returns to scale, input oriented model vary from the other models by more than one percent. In fact, DMU H31(a 2002 Medical Center) is nearly 15% higher under the EMS formulation. These EMS values are not believed to be correct.

Another data set was selected and analyzed with the assumption that these findings were spurious. This analysis resulted in nearly identical findings. The results of the EMS analysis, therefore, are suspect. All models in this dissertation are therefore formulated in GAMS and confirmed in DEA Solver.

Appendix 2-One Phase GAMS Code Sample for DEA Model, Remarked

```
$title Comprehensive Data Envelopment Analysis-Larry Fulton
$ontext

One phase solution to the BCC-I model
Loosely adapted from the GAMS DYSON DEA model.

$offtext
$offlisting

*Establish sets and subsets for the DMUs, inputs, and outputs
sets          d    dmus
              ds(d) selected dmus
              j    inputs and outputs
              i(j)  inputs only
              o(j)  outputs only
              k(j)  nondiscretionary inputs;

*Parameters are used to specify known values (e.g., sample values and bounds)
Parameter       data(d,j) unit  input  output
              vlo      v lower bound
              ulo      u lower bound
              norm     normalizing constant;

*Variables are manipulated by the solver.
Variables       v(i)    input weights
              u(o)    output weights
              eff     efficiency CRS
              convex  dual convexity for VRS
              dsn(i)  dual input slack  (neg)
              dsn2(k) non-discretionary input slack
              dsp(o)  dual output slack (pos)
              lambda(d) multiplier for VRS
              z       primal obj function for VRS

*Some variables are specified as strictly non-negative.
positive variables z,dsn(i),dsp(o),lambda(d);

*Equations are specified distinctly for inclusion in the model.
Equations
        primobj       primal objective funtion
        slackneg(d,i)  input excesses
        slackpos(d,o)  output shortages
        convexity      VRS
        nondis(d,k)    nondiscretionary input  ;

*In the next section, the equations are flushed out with variables and parameters.
*Dual BCC-Input Oriented Model (VRS)
*One should note that this set of equations is solved repeatedly for EACH DMU
```

```
*Obj fx:  Efficiency = 100*z – non-Archimedean element*(sum of input excesses minus output
shortages)
primobj..        eff =e= norm*z - vlo*sum(i, dsn(i)) - ulo*sum(o, dsp(o));
*Constraint:  dual multipliers*input data + dual input slack = efficiency * input data for selected DMU
slackneg(ds,i).. sum(d, lambda(d)*data(d,i)) + dsn(i) =e= z*data(ds,i);
*Constraint:  dual multipliers*output data - dual output slack = output data for the selected DMU.
slackpos(ds,o).. sum(d, lambda(d)*data(d,o)) - dsp(o)=e= data(ds,o);
*see Cooper, pg. 64 for the inclusion of slack in this formulation
nondis(ds,k)..   sum(d, lambda(d)*data(d,k)) + dsn2(k)=e=data(ds,k);
*The convexity constraint ensures a VRS model as opposed to the CRS.
convexity..      sum(d, lambda(d))=e=1;

model            primvrs primal /primobj, slackneg, slackpos, nondis, convexity/


*Assign variables and DMUs to the sets.
sets           d       units / H1*H72/
             j       inputs and outputs /POP, ENR, TC, RWP, RVU,PREV, ACC/
             i(j)    inputs            /ENR, TC/
             k(j)    nondiscretionary   /POP/
             o(j)    outputs /  RWP, RVU, PREV, ACC  /


*The sample data follows.
Table data(d,j)
         POP        ENR        TC          RWP        RVU          PREV        ACC
H1        22.08       14.55       49.16        8.02        119.46        87.61       2.99
H2        28.96       22.18       64.23        9.19        229.15        87.79       1.49
TRUNCATED

$eolcom //
;
*Here,  assign values to the parameters in the equations (non-Archimedean element and normalization
constant).
vlo=1e-4;
ulo=1e-4;
norm=1;

*Here, establish a set consisting of all the DMUs for analysis.
set ii(d) /h1*h72/

*Now, prepare to make a report, which has slack values and efficiency scores.
parameter rep Summary Report;
*Here, loop through all of the DMUs, solving the model formulation for each one.
loop (ii, ds(ii)=yes;
solve primvrs using lp min eff;
*Here is the report format and values.
rep(ii, 'Obj Fx=') = primvrs.objval;
rep(ii, 'Eff =')=z.l;
rep(ii, 'output slack') = sum(o,dsp.l(o));
```

```
rep(ii, 'input slack')=sum(i,dsn.l(i));
rep(ii, 'ENROLL slack') = dsn.l('ENR');
rep(ii, '$ slack') = dsn.l('TC');
rep(ii, 'RWP/000 slack') = dsp.l('RWP');
rep(ii, 'RVU/0,000 slack') = dsp.l('RVU');
rep(ii, 'PREV slack') = dsp.l('PREV');
rep(ii, 'ACC slack') = dsp.l('ACC');
ds(ii)=no);
display rep;
```

Appendix 3-Changes in GAMS Coding from One Phase to Two Phase

Phase 1:

```
*The objective function reduces to a single variable.
primobj..       eff =e= norm*z;
*Slacks are removed from the constraints.
slackneg(ds,i).. sum(d, lambda(d)*data(d,i)) =l= z*data(ds,i);
slackpos(ds,o).. sum(d, lambda(d)*data(d,o)) =g= data(ds,o);
*Nondiscretionary variables are forced to equality.
nondis(ds,k)..   sum(d, lambda(d)*data(d,k)) =e=data(ds,k);
*The convexity constraint remains.
convexity..      sum(d, lambda(d))=e=1;
```

Appendix 4-Comparison of Technical Efficiency Scores

The outlined items reflect the discrepancies between the efficiency scores generated by the EMS solver.

DMU	EMS	DEA Solver	1-Phase GAMS	2-Phase GAMS
H1	100.00%	100.00%	100.00%	100.00%
H2	100.00%	95.08%	95.00%	95.10%
H3	80.46%	80.46%	80.40%	80.50%
H4	100.00%	100.00%	100.00%	100.00%
H5	96.98%	96.22%	96.10%	96.20%
H6	100.00%	100.00%	100.00%	100.00%
H7	100.00%	100.00%	100.00%	100.00%
H8	79.63%	79.61%	79.60%	79.60%
H9	67.48%	67.49%	67.40%	67.50%
H10	100.00%	100.00%	100.00%	100.00%
H11	98.04%	98.08%	97.90%	98.00%
H12	100.00%	100.00%	100.00%	100.00%
H13	100.00%	100.00%	100.00%	100.00%
H14	100.00%	100.00%	100.00%	100.00%
H15	80.02%	80.01%	80.00%	80.00%
H16	85.71%	85.71%	85.60%	85.70%
H17	88.26%	88.29%	88.20%	88.30%
H18	96.16%	96.12%	96.10%	96.20%
H19	94.30%	94.30%	94.30%	94.30%
H20	100.00%	100.00%	100.00%	100.00%
H21	100.00%	100.00%	100.00%	100.00%
H22	100.00%	100.00%	100.00%	100.00%
H23	81.76%	81.75%	81.60%	81.80%
H24	100.00%	100.00%	100.00%	100.00%
H25	97.63%	97.66%	97.60%	97.60%
H26	100.00%	100.00%	100.00%	100.00%
H27	75.71%	75.65%	75.60%	75.70%
H28	100.00%	100.00%	100.00%	100.00%
H29	100.00%	100.00%	100.00%	100.00%
H30	92.36%	92.36%	92.00%	92.40%
H31	95.75%	81.28%	81.20%	81.30%
H32	82.63%	82.52%	82.60%	82.60%
H33	67.11%	67.12%	67.10%	67.10%
H34	91.42%	91.42%	91.40%	91.40%
H35	77.58%	77.59%	77.60%	77.60%
H36	100.00%	100.00%	100.00%	100.00%

H37	100.00%	100.00%	100.00%	100.00%
H38	100.00%	100.00%	100.00%	100.00%
H39	75.12%	75.10%	75.10%	75.10%
H40	73.43%	73.44%	73.30%	73.40%
H41	95.77%	95.69%	95.70%	95.80%
H42	100.00%	100.00%	100.00%	100.00%
H43	92.31%	92.40%	92.20%	92.30%
H44	100.00%	100.00%	100.00%	100.00%
H45	100.00%	100.00%	100.00%	100.00%
H46	100.00%	100.00%	100.00%	100.00%
H47	75.17%	75.16%	75.00%	75.20%
H48	91.54%	91.54%	91.40%	91.50%
H49	100.00%	92.76%	92.70%	92.70%
H50	78.90%	77.75%	77.70%	77.80%
H51	57.64%	57.63%	57.40%	57.60%
H52	100.00%	100.00%	100.00%	100.00%
H53	88.31%	88.27%	88.10%	88.30%
H54	79.56%	79.55%	79.30%	79.60%
H55	75.40%	68.49%	68.50%	68.50%
H56	66.42%	66.38%	66.30%	66.40%
H57	49.89%	49.90%	49.80%	49.90%
H58	86.71%	86.68%	86.50%	86.60%
H59	100.00%	100.00%	100.00%	100.00%
H60	86.82%	84.73%	84.70%	84.70%
H61	80.98%	80.98%	80.90%	81.00%
H62	87.42%	87.42%	87.30%	87.40%
H63	78.68%	78.73%	78.50%	78.70%
H64	67.79%	67.79%	67.60%	67.80%
H65	100.00%	100.00%	100.00%	100.00%
H66	91.84%	91.74%	91.70%	91.80%
H67	82.94%	83.56%	81.90%	82.90%
H68	91.32%	91.32%	91.30%	91.30%
H69	100.00%	98.15%	98.10%	98.10%
H70	80.00%	80.00%	79.40%	80.00%
H71	67.37%	67.37%	67.10%	67.40%
H72	75.00%	75.00%	74.90%	75.00%

Appendix 5-Sampling Methods

Inverse CDF Sampling

The most basic method for drawing samples from a multivariate distribution is probably inversion of the Cumulative Distribution Function (CDF); however, this method is not available for complex distributions where the closed form version is unobtainable. Nevertheless, it is important to discuss this transformation as it serves a role in sampling from multivariate distributions. Given a CDF $F(x)$, the distribution of y is always known: $y \sim U(0, 1)$, with density

$$f(y \mid a=0, b=1) = \frac{1}{b-a} = 1.$$

This fact derives from the definition of a probability distribution, which exists on the domain [0, 1], and it yields an algorithm for producing any individual x value based upon the value of y. The algorithm is readily illustrated in the univariate case. Assume that the random variable X is distributed as an exponential, e.g., $X \sim exp\left(\frac{1}{\lambda}\right)$. Then

$$f(x) = \lambda e^{-\lambda x} I_x(0, \infty)$$

and the CDF is stated in closed form as

$$F(x) = 1 - \lambda e^{-\lambda x} I_x(0, \infty).$$

With the CDF in hand, the generic algorithm is simple (*Introduction to Monte Carlo Methods,* 1996, pg. 25).

1. Draw a sample from $y \sim U(0,1)$.
2. Equate $y=F(x)$.
3. Solve for $x=F^{-1}(y)$

If λ=1 and one draws ξ from the distribution of y, then $x = - \log(1 - \xi)$. The chart below illustrates the transformation.

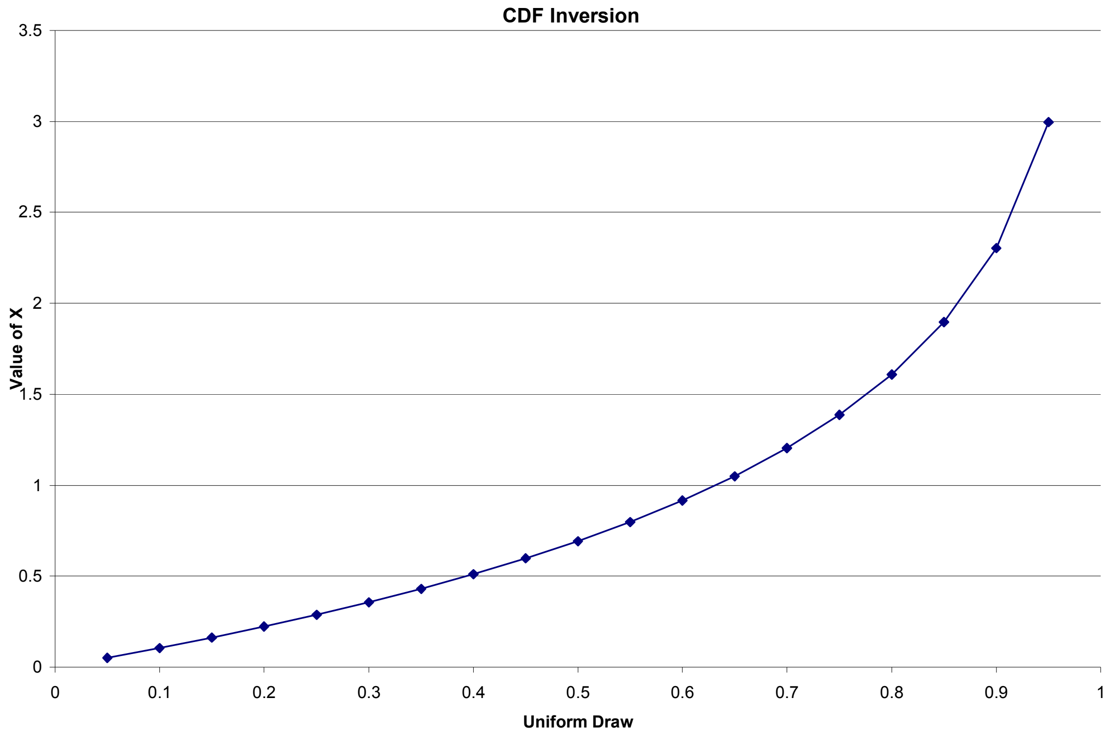


Figure 36. CDF Inversion.

The CDF inversion results in a draw from a CDF with closed form. In this case, draws result in an exponential distribution.

The CDF transform is useful for only the most basic distributions and has other limitations as well. In general, complex distributions require other methods.

Acceptance-Rejection Sampling

One method for drawing samples from multivariate distributions is the Acceptance-Rejection (AR), which requires only knowledge of the joint probability density or mass function. The researcher seeks to generate samples from the continuous target density $\pi(x) = \frac{f(x)}{K}, x \in R^d$. In this case, K is a normalizing constant, which may be unknown. Allow $g(x)$ to be another density that may be simulated by some known method (e.g., CDF inversion). Allow c to be a constant such that $f(x) \leq cg(x)$ $\forall x$. Then the algorithm is succinctly presented below:

1. Generate a draw z from the $g(.)$ density.
2. Generate a draw u from a $U \sim U(0,1)$ density.
3. If $u \leq f(z) / cg(z)$ then keep z else step 1

This method requires the statistician to find a constant c and a density $g(x)$ that dominate $f(x)$. If the multivariate distribution is extremely complicated, the dominating function may be so large that the number of rejections is extremely high. The expected number of iterations needed to complete a single sample is *1/c* so *c* is optimized by $c = -\sup_x \frac{f(x)}{g(x)}$ (Chib and Greenberg, 1995).

For a simple example, assume that *n=100* observations are distributed *X~Bern(p)* and the prior on *p~Gamma(α,β)* with $\alpha=\beta=1$. The density for the gamma is expressed

$$f(p) = \frac{p^{\alpha-1}}{\Gamma(\alpha)\beta^{\alpha}} \exp\left(\frac{-p}{\beta}\right) I_p(0,\infty).$$

Given *X~Bern(p)* and the prior on *p~Gamma(α,β)* with $\alpha=\beta=1$, the posterior distribution is easy to derive. The derivation follows immediately.

$$\pi(p \mid x) \propto f(T(x) \mid p) * f_{prior}(p) =$$

$$\binom{n}{x} p^{x}(1-p)^{n-x} \; \frac{p^{\alpha-1} e^{-\frac{p}{b}}}{\Gamma(\alpha)\mathrm{B}^{\alpha}} =$$

$$\binom{n}{x} p^{x+\alpha-1}(1-p)^{n-x} \; \frac{e^{-\frac{p}{b}}}{\Gamma(\alpha)\mathrm{B}^{\alpha}} \propto$$

$$p^{x+\alpha-1}(1-p)^{n-x} exp(-p / \beta) =$$

$$p^{x+2}(1-p)^{100-x} exp(-p)$$

The kernel of the density may not be readily recognizable, but a graph is quite revealing. Assume fixed $x = 50$, and the graph of the distribution appears in the following figure.

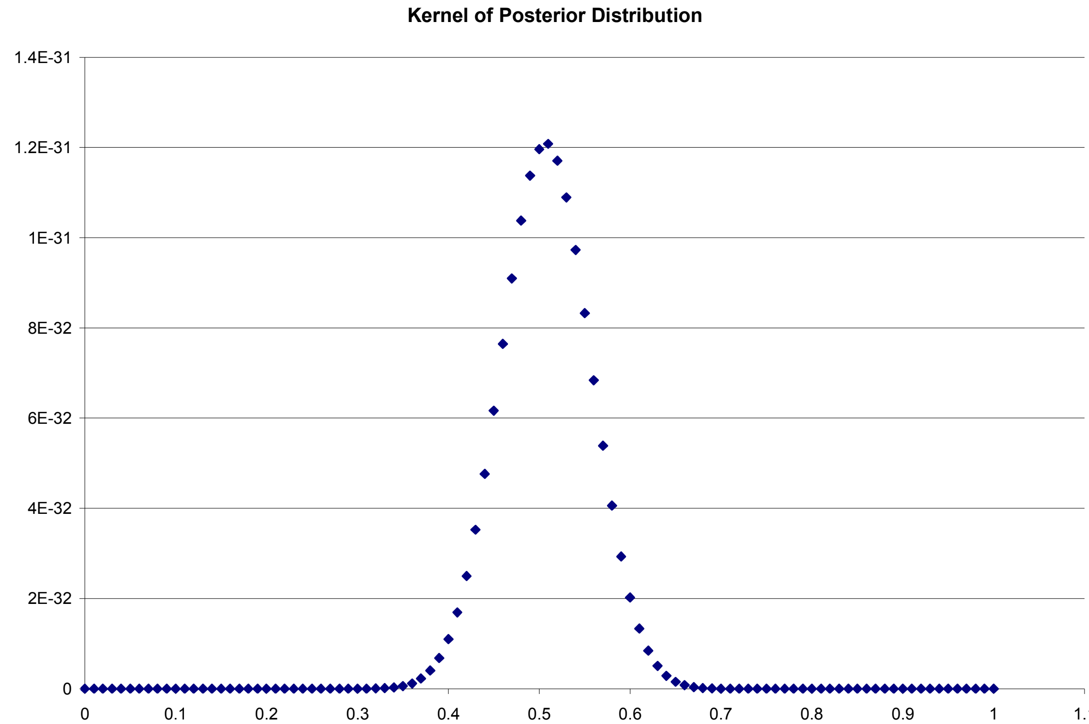


Figure 37. Graph of Posterior Distribution.

The graph of the kernel of the posterior distribution appears to be a symmetrical distribution similar to a Beta distribution as it is finite on (0,1).

Finding a dominating function is not difficult in this case. Because of the distribution, one could select a symmetric Beta, t, normal, or other distribution. A convenient selection is to use the normal distribution. Since the posterior is symmetric, selection of a normal distribution with the same mean as the posterior leaves only the variance and the value of c for determination. Through a quick error minimization algorithm (omitted), one can determine that the dominating density $g \sim N(.51, .0064)$ with c = 2.42183E-32 is one of many that would work. The following illustration shows the domination and the minimal error associated with the choice.

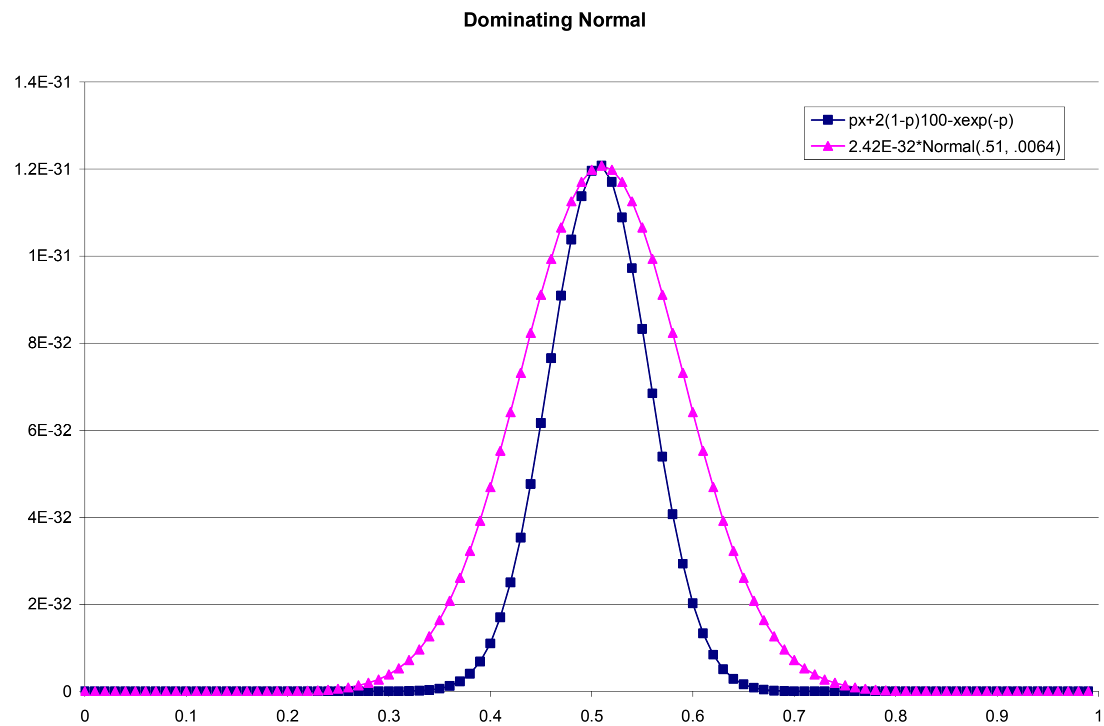


Figure 38. Dominating Density Function.

From this figure, it becomes clear that a dominating density has been found. This density will serve in conjunction with a uniform density in the Acceptance-Rejection Algorithm.

To finish the example for this algorithm, one generates a draw from the normal density. Conveniently, use of the Box-Muller algorithm generates two simultaneous draws from two separate independent standard normal distributions, and the location-scale transform serves to translate these draws. A discussion of the algorithm follows.

Assume that $x_1 \sim U(0,1)$, $x_2 \sim U(0,1)$. Then one can prove that y_1, y_2 are independent, standard, normal variables using the following transformations.

1. Draw x_1 and x_2.
2. Calculate $y_1 = cos(2\pi x_1) * (-2ln(x_2))^{1/2}$
3. Calculate $y_2 = \sin(2\pi x_1) * (-2\ln(x_2))^{1/2}$

One can readily show that this transformation yields two independent standard normal variables. Since x_1 and x_2 are uniform variables on (0,1), the transform $a = 2\pi x_1$ is uniform on $(0, 2\pi)$ as illustrated: $f(x_1)= I_x(0, 1)$, $a = 2\pi x_1$ for $a \in (0, 2\pi)$, $g^{-1}(x_1) = a / 2\pi$. Since this tranformation is monotonic throughout and continuous derivatives exist for g^{-1}, a simple transformation applies:

$$f(a) = f(g^{-1}(x_1)) * \left| \frac{d}{dg^{-1}(x_1)} \right| \text{ for } a \in A.$$

$$\left| \frac{d}{dg^{-1}(x_1)} \right| = 1/2\pi,\ f(a) = 1/2\pi,\ a \in (0,2\pi) \Rightarrow a \sim U(0, 2\pi).$$

Further, if $b = -2ln(x_2)$, then $b \sim exp(1/2)$, and the 2d transformation is trivial as well:

$$f(x_2) = I_x(0, 1),\ b = -2ln(x_2) \text{ for } b \in (0, \infty),\ g^{-1}(b) = exp(-b/2).$$

The transformation is monotonic (decreasing from zero to one) and continuous derivatives exist. The simple transformation

$$f(b) = f((g^{-1}(x_2)) * \left| \frac{d}{dg^{-1}(x_2)} \right| \text{ for } b \in B$$

applies. Proceeding with this transformation, one generates the following:

$$\left| \frac{d}{dg^{-1}(x_2)} \right| = \left| \frac{-1}{2} * \left(\exp\left(\frac{-b}{2} \right) \right) \right| \text{ so } f(b) = \frac{1}{2} * \left(\exp\left(\frac{-b}{2} \right) \right) \text{ for b} \in (0, \infty).$$

Immediately, one can see that $b \sim exp(2)$.

With these two preliminary transformations, the joint distribution of a and b may be expressed as

$$f(a,b) = \frac{1}{4\pi} \exp\left(\frac{-b}{2} \right) \text{ for } a \in (0, 2\pi),\ b \in (0, \infty).$$

Now, the transformations y_1 and y_2 are readily expressed as follows:

$$y_1 = (b)^{1/2} * cos(a) \text{ and } y_2 = (b)^{1/2} * sin(a).$$

From the joint distribution of $f(a,b)$, it is apparent that the variable b may be expressed readily in terms of y_1 and y_2. Squaring both y_1 and y_2 and adding them together yields the obvious solution

$$y_1^2 + y_2^2 = b * cos^2(a) + b * sin^2(a) = b.$$

Further, the relationship for a may now be expressed readily by dividing y_2 by y_1:

$$(y_2 / y_1) = tan(a) \text{ so } a = tan^{-1}(y_2 / y_1).$$

The remainder of the transformation follows.

$$a = tan^{-1}(y_2 / y_1),\ b = y_1^2 + y_2^2.$$

Under certain regularity conditions (see Casella and Berger), the transformation

$$f(y_1,y_2) = f((g^{-1}(x_1),\ (g^{-1}(x_2))*\left\|J\right\|,$$

where $\left\|J\right\|$ is the absolute value of the determinant of *the* Jacobian matrix defined as

$$\begin{vmatrix} \dfrac{\partial y_1}{\partial a} & \dfrac{\partial y_2}{\partial a} \\ \dfrac{\partial y_1}{\partial b} & \dfrac{\partial y_2}{\partial b} \end{vmatrix}.$$

produces the joint distribution f(a,b). Furthermore, it may shown that

$$\left\|J^{-1}\right|^{-1}\Bigg| = \left\|\begin{matrix} \dfrac{\partial a}{\partial y_1} & \dfrac{\partial a}{\partial y_2} \\ \dfrac{\partial b}{\partial y_1} & \dfrac{\partial a}{\partial y_2} \end{matrix}\right|^{-1}\Bigg|$$

produces the same results (see Ross). Proceeding with the transformation,

$$f(y_1,y_2) = f((g^{-1}(x_1),\ (g^{-1}(x_2))*\ \left\|J\right\| = f(y_1,y_2) = \frac{1}{4\pi}\exp\left(\frac{-\left(\mathrm{y_1}^2 + \mathrm{y_2}^2\right)}{2}\right) * \left\|J^{-1}\right|^{-1}\Big|.$$

Next, one calculates the following.

$$\left\|J^{-1}\right|^{-1}\Big| = \left\|\begin{matrix} \dfrac{\partial a}{\partial y_1} & \dfrac{\partial a}{\partial y_2} \\ \dfrac{\partial b}{\partial y_1} & \dfrac{\partial a}{\partial y_2} \end{matrix}\right|^{-1}\Bigg| = \left\|\begin{matrix} -\sin(a)*\sqrt{b} & \dfrac{\cos(a)}{2\sqrt{b}} \\ \cos(a)*\sqrt{b} & \dfrac{\sin(a)}{2\sqrt{b}} \end{matrix}\right|^{-1}\Bigg| =$$

$$\left|\left(\frac{-\sin^2(a)}{2} - \frac{\cos^2(a)}{2}\right)^{-1}\right| = \left|\frac{-1}{2}\right|^{-1} = 2.$$

So $f(y_1,y_2) = \frac{1}{4\pi}\exp\left(\frac{-\left(y_1^2+y_2^2\right)}{2}\right)*2 =. \frac{1}{2\pi}\exp\left(\frac{-\left(y_1^2+y_2^2\right)}{2}\right)$.

Note that $y_1 \in (-\infty, \infty)$ and $y_2 \in (-\infty, \infty)$ (no domain dependence), so separation is simple. The marginals are readily expressed as

$f(y_1) = \frac{1}{\sqrt{2\pi}}\exp\left(\frac{-\left(y_1^2\right)}{2}\right)$ *for* $y_1 \in (-\infty, \infty)$ *and* $\frac{1}{\sqrt{2\pi}}\exp\left(\frac{-\left(y_2^2\right)}{2}\right)$ *for* $y_2 \in (-\infty, \infty)$,

which are independent standard normal variables.

A second method for generating the draws from two independent standard normals follows.
Given $x_i \sim U(-1,1)$, let $R = x_1^{\ 2} + x_2^2$. If $R = 0$ or $R > 1$, resample.
$z_1 = x_1 * (-2\ln(R)/R)^{1/2}$
$z_2 = x_2 * (-2\ln(R)/R)^{1/2}$
While the first method discussed distributes draws within the unit circles using polar coordinates, the second method uses cartesian coordinates to accomplish the same feat using an AR algorithm. The second method is usually faster since it employs only one transcendental function, although it throws away $4/\pi$-1 or about 27 percent of the sampled pairs. (http://en.wikipedia.org/wiki/Box-Muller_transformation). The example for this algorithm is omitted.
For the specific example discussed, the full algorithm follows.

1. Generate two draws from two separate Uniform (0,1) densities.
2. Determine which z_i variable will represent the normal distribution for the problem.
3. Calculate $z_{i.}$ This value is the draw from the *g(.)* density
4. Generate a draw *u* from a Uniform(0,1) density.
5. If $u \leq f(z) / cg(z)$ then keep *z* else step 1.

The AR algorithm works well when the dominating function can be found. For multi-modal distributions, discovery of a workable *g(.)* may be non-trivial.
Metropolois-Hastings Algorithm (Gibbs Sampling)
A third method for generating values from a multivariate distribution is the Metropolis-Hastings (M-H) algorithm, a Markov Chain Monte Carlo (MCMC) method. Markov Chains are typically concerned with finding the invariant distribution, whereas MCMC methods are concerned with discovering the transition kernel from the target density. With the specification of the multivariate likelihood function, the researcher has obtained the target density. Finding the transition kernel is the focus, hence the use of the M-H algorithm.

The M-H algorithm uses a candidate generating density (say Q) for consideration of movement from one state to another. To illustrate the algorithm concretely, assume that the probability of moving from one state to another state is assumed to be Multivariate Normal (MVN), specifically

$$Q(x_{candidate}, x_t) \sim MVN(x_{candidate} - x_t, \Sigma).$$

Using this proposal density, a candidate is drawn with probability $Q(x_{candidate}, x_t)$ and then the value $\alpha = \alpha 1 * \alpha 2$ is calculated where

$$\alpha 1 = \pi(x_{candidate})/\pi(x_t) \text{ and } \alpha 2 = Q(x_t, x_{candidate}) / Q(x_{candidate}, x_t).$$

The value $\alpha 1$ is the likelihood ratio for the proposed sample and the current sample, while $\alpha 2$ is the likelihood ratio for the movement from the current state to the next state versus from the next state to the current state. If the proposal density is symmetric, then this value is always one. Once α is calculated, the next state space is determined as follows:

$$x_{t+1} = \begin{cases} x_{candidate} \text{ if } \alpha \geq 1 \\ x_{candidate} \text{ with probability } \alpha \text{ if } \alpha < 1 \\ x_t \text{ with probability } 1 - \alpha \text{ if } \alpha < 1 \end{cases}$$

To start the algorithm, initial values x_0 are provided or generated and the algorithm is run several thousand times for burn-in (the procedure whereby initial values are forgotten, e.g., shuffling the cards.). The algorithm is succinctly stated as follows:

1. Generate $x_{candidate}$ from Q. Generate u from U(0, 1).
2. Calculate $\alpha = \alpha 1 * \alpha 2$.
3. If $\alpha \geq u$, *then* $x_{t+1} = x_{candidate}$ Else $x_{t+1} = x_t$ and go to step 1.
4. Return all values for x.

This algorithm derives from Chib and Greenberg's (1995) algorithm, which is arguably one of the best algorithms in print.

The problem with the M-H algorithm is that an appropriate proposal density must be found. The random walk proposal of Metropolis et al (1953) was used for the example above with an MVN distribution. The probability for movement from x_t to $x_{candidate}$ is $Q(x_{candidate} - x_t)$, which is equates to the current value plus noise. The choice of the multivariate density Q is again an issue. Another choice for the movement from x_t to $x_{candidate}$ is the selection of an independence chain, $Q(x_{candidate})$ (Hastings, 1970); however, both location and spread must be selected, whereas in the random walk the location is partially determined. Chib and Greenberg (1994) suggest a modification. If the probability distribution $\pi(t)$ may be written as $\pi(t) \propto h(t) * \Psi(t)$, with $\Psi(t)$ uniformly bounded, then select $Q(x_{candidate}, x_t) = h(x_{candidate})$ and evaluate the draw with $\alpha = min\{\Psi(x_{candidate}) / \Psi(x_t), 1\}$. A fourth method is to use the A-R method with a pseudo-dominating density (Tierney, 1994). A fifth method involves autoregressive chains. $Q(x_{candidate}, x_t) = Q(x_{candidate} - a - B(x - a)$ where a is a vector, B is a matrix set equal to –

I. This formula derives from the linear formula $y = a + B(x-a) + z$, where z has density Q (Tierney, 1994).

Assuming that an appropriate density is found (e.g., Multivariate Normal), the next question of importance is determining the spread. For random walks, acceptance rate calculations may help adjust the spread (Roberts, Gelman, and Gilks, 1994). For independent proposal densities, the tails should dominate (Chib and Greenburg, 1994). Autocorrelation analysis of the variables provides insight into the effect of the spread as well. In some circumstances, autocorrelation will remain high when the spread is adjusted appropriately. In these cases, different candidate densities must be considered.

The M-H algorithm is the basis for Gibbs Sampling. By sampling repeatedly from the conditional distribution of the posterior, the marginal distributions of the desired variables emerge. This fundamental principle may be stated in the two variable case as follows (Casella and George, 1992)

$$X'_j \sim f(x \mid Y'_j = y'_j)$$
$$Y'_{j+1} \sim f(y \mid X'_j = x'_j)$$

Once the initial value y_0 is specified, the value for x_o is determined. This value is then used to determine y_1. The beauty in this technique is that large enough samples from the conditional distributions are as is if one was sampling from the marginal distributions (Gelfand and Smith, 1990).

Auxiliary Variable Technique

One of the most elegant methods for sampling from a conditional distribution involves the introduction of auxiliary variables, which are distributed uniformly. Assume that one wishes to sample such a distribution. Assume that this distribution may be expressed $f(\mu) \propto \pi(\mu)*a(\mu)*b(\mu)*c(\mu)$, where π is a density of known form and the functions *a, b,* and *c* are non-negative invertible functions. Note: these three functions need not be densities. Under these assumptions, the distribution of f may be sampled and all but one of the full conditionals are distributed as uniform random variables with the remaining conditional a truncated version of π (Damien, Wakefield, Walker, 1999).

To illustrate this theorem, introduce latent variables *Ua, Ub,* and *Uc* defined on $(0,\infty)$ such that the joint density with μ is expressed

$$f(\mu, ua, ub, uc) \propto \pi(\mu)* I_{(ua)}(0, a(\mu))*I_{(ub)}(0, b(\mu))*I_{(uc)}(0, c(\mu)).$$

The marginal of μ is $f(\mu)$ and the conditionals for *Ua, Ub,* and *Uc* are clearly uniform truncated by their respective functions. The full conditional for μ is $\pi(\mu)$ restricted to the set $A = [\mu: a(\mu) > ua, b(\mu) > ub, c(\mu) > uc]$. The beauty in this approach is that the selection of variables U need not be restricted to uniform random variables (Damien, Wakefield, Walker, 1999).

Comparing Bayesian Models.

One method for comparing Bayesian models involves computation of the posterior odds using the Bayes' Factor and the priors. Bayes' Factor is defined as the likelihood ratio for competing models. Assume that one is testing two specified models (*Ho:* Λ versus *Ha:* Ψ). The mathematical expression for the Bayes' Factor (BF) becomes BF = $\frac{P(X \mid \Lambda)}{P(X \mid \psi)}$, which is obviously a likelihood ratio. To reflect the ability to vary prior beliefs, the posterior odds equation, $\frac{P(\Lambda \mid X)}{P(\psi \mid X)}$, is derived as $\frac{P(\Lambda \mid X)}{P(\psi \mid X)} = \frac{P(X / \Lambda) * P(\Lambda)}{P(X / \psi) * P(\Psi)}$. The posterior odds are then evaluated against a somewhat arbitrary table for determining the better model (see following table). One notices immediately that if the priors are equally likely (e.g., $P(\Lambda) = P(\Psi)$), then the equation is nothing more than the Bayes' Factor itself.

Table 54. Bayes' Factor.

The following table is widely used for acceptance or rejection of the null hypothesis using Bayes' Factor methodology (Jeffreys, 1961).

Posterior Odds Ratio	*Model Determination*
$\frac{P(\Lambda \mid X)}{P(\psi \mid X)} > 1$	Null hypothesis supported
$\frac{P(\Lambda \mid X)}{P(\psi \mid X)} < 1$	Barely worth a mention
$\frac{P(\Lambda \mid X)}{P(\psi \mid X)} < \frac{1}{10}$	Evidence against the null is substantial
$\frac{P(\Lambda \mid X)}{P(\psi \mid X)} < \frac{1}{\sqrt{10}}$	Strong evidence
$\frac{P(\Lambda \mid X)}{P(\psi \mid X)} < \frac{1}{10^{3/2}}$	Very strong evidence
$\frac{P(\Lambda \mid X)}{P(\psi \mid X)} < \frac{1}{100}$	Decisive evidence

Tracking the Deviance Information Criterion (DIC), which approximates the expected posterior loss, provides an alternate method for determining which Bayesian model appears to fit the data best (Spiegelhalter D J, Best N G, Carlin B P and van der Linde A, 2002). The algorithm (submitted without proof) follows.

The posterior mean of the deviance ($\overline{D}$) equals $E_{\beta|Z}(-2Log(L(Z \mid \beta)))$. The variable Z includes all stochastic nodes that have data assigned. Substituting the posterior means of the parameters into $-2Log(L(Z \mid \beta))$ produces the point estimate of the deviance (at time t) or $\hat{D} = E_{\beta|Z}(-2Log(L(Z \mid \overline{\beta})))$. The effective number of parameters in play is estimated as $\left|\overline{D} - \hat{D}\right|$ and is equivalent to a penalty term for increasing model complexity. This term may also be construed to indicate the reduction in uncertainty resulting from estimation. Finally, the deviance information criterion becomes DIC = $\hat{D}$ - 2*($\left|\overline{D} - \hat{D}\right|$). The model with the lowest DIC is selected as the best fit model.

Appendix 6-RAM Formulation (2001 Data)

```
$title Comprehensive Data Envelopment Analysis-Larry Fulton
$ontext

Adapted from the GAMS DYSON DEA model..

$offtext
$offlisting

sets        d    dmus
            ds(d) selected dmus
            j    inputs and outputs
            i(j)  inputs only
            o(j)  outputs only
            k(j)  nondiscretionary inputs;
*           l(j)  nondiscretionary outputs

Parameter     data(d,j) unit  input  output
            vlo      v lower bound
            ulo      u lower bound
            norm     normalizing constant;

Variables     v(i)    input weights
            u(o)    output weights
            eff     efficiency
            convex  dual convexity for VRS
            dsn(i)  dual input slack  (neg)
            dsn2(k) non-discretionary input slack
            dsp(o)  dual output slack (pos)
            lambda(d) multiplier for VRS
           ;

positive variables z,dsn(i),dsp(o),lambda(d);

Equations
        primobj       primal objective funtion
        slackneg(d,i)  input excesses
        slackpos(d,o)  output shortages
        convexity      VRS
        nondis(d,k)    nondiscretionary input ;
*nondis2(d,l)   nondiscretionary output
```

```
*RAM Model
*Obj fx:  maximizing the slacks over the ranges
primobj..        eff =e= sum(i, dsn(i)/(smax(ds, data(ds,i))-smin(ds,data(ds,i))+.001)) +
sum(o, dsp(o)/(smax(ds, data(ds,o))-smin(ds,data(ds,o))+.001));
*Constraint:  dual multipliers*input data + dual input slack = efficiency * input data for
selected DMU
slackneg(ds,i).. sum(d, lambda(d)*data(d,i)) + dsn(i) =e= data(ds,i);
slackpos(ds,o).. sum(d, lambda(d)*data(d,o)) - dsp(o)=e= data(ds,o);
*nondiscretionary variable
nondis(ds,k)..   sum(d, lambda(d)*data(d,k)) =e=data(ds,k);
*nondis2(ds,l)..  sum(d, lambda(d)*data(d,l)) =e=data(ds,l);
convexity..      sum(d, lambda(d))=e=1;

model           primvrs primal /primobj, slackneg, slackpos, nondis, convexity/

*The sets are currently defined to reflect cost as an input (budget flowing in)
*and those served coming out (Active Duty dependents, Retirees, etc.)
*This is the global model for all catchment areas around the major facilities.

sets            d       units / H1*H24/
                j       inputs and outputs /Pop, Total, Work, Qual, Access/
                i(j)    inputs            /Total,Pop/
                k(j)    nondiscretionary   /Pop/
                o(j)    outputs /  Work, Qual, Access /


*Notes:  This is 2001-2003 data.

Table data(d,j)
        Pop        Total       Work        Qual        Access
H1      1.012       3.895       1.422       4.000       4.280
H2      1.524       4.162       1.782       4.000       4.190
H3      2.773       4.910       2.353       4.000       4.140
H4      2.628       5.844       3.179       5.000       4.240
H5      3.241       5.452       2.843       4.000       4.120
H6      3.098       5.079       2.383       5.000       4.120
H7      2.426       5.495       2.646       5.000       4.300
H8      2.419       4.769       2.188       4.000       4.200
H9      2.064       4.687       1.825       4.000       4.120
H10      1.706       4.054       1.735       5.000       4.180
H11      1.061       4.265       1.370       4.000       4.190
```

```
H12     1.558   4.478   2.014   5.000   4.260
H13     1.239   3.921   1.258   5.000   4.230
H14     3.120   5.670   3.090   4.000   4.310
H15     2.470   4.733   2.171   4.000   4.230
H16     2.076   4.193   1.573   4.000   4.190
H17     1.936   4.454   1.805   5.000   4.260
H18     2.070   4.376   2.047   4.000   4.170
H19     2.892   5.819   3.108   4.000   4.210
H20     2.152   6.383   3.358   6.000   4.130
H21     0.419   3.557   1.103   4.000   4.150
H22     2.221   5.136   2.485   5.000   4.180
H23     2.527   4.559   2.107   4.000   4.070
H24     3.293   5.482   2.987   5.000   4.110

$eolcom //


set ii(d) /h1*h24/
*Use this next line for CRS model
*convex.fx = 0;


parameter rep Summary Report;

loop (ii, ds(ii)=yes;


solve primvrs using lp max eff;
*rep(ii, 'Obj Fx=') = primvrs.objval;
rep(ii, 'output slack') = sum(o,dsp.l(o));
rep(ii, 'input slack')=sum(i,dsn.l(i));
rep(ii, '$ slack') = dsn.l('Total');
rep(ii, 'Work slack') = dsp.l('Work');
rep(ii, 'Qual slack') = dsp.l('Qual');
rep(ii, 'ACC slack') = dsp.l('Access');


ds(ii)=no);
display rep;
```

Appendix 7-WINBUGS Code

```
Model{

for(i in 1:24){

y[i]~dnorm(mu[i],tau)I(0,)

#tau is the precision, not the variance

z1[i]<-(pop[i]-mean(pop[]))/sd(pop[])
z2[i]<-(work[i]-mean(work[]))/sd(work[])
z3[i]<-(qual[i]-mean(qual[]))/sd(qual[])
z4[i]<-(access[i]-mean(access[]))/sd(access[])
z5[i]<-(eff[i]-mean(eff[]))/sd(eff[])
z6[i]<-(medcen[i]-mean(medcen[]))/sd(medcen[])
z7[i]<-(medwork[i]-mean(medwork[]))/sd(medwork[])


mu[i] <- b[1]*z1[i] + b[2]*z2[i] + b[3]*z3[i]+ b[4]*z4[i]+ b[5]*z5[i] +b[6]*z6[i]+b[7]*z7[i]

#Set markers for observed bias and error ESS

}

#All precisions are considered vague

tau~dgamma(.0001,.0001)
sigma<-1/sqrt(tau)

#Model 1:  The parameters are considered to be independent.
#for(j in 1:7){b[j]~dnorm(0,.00001)}

#Model 2:  The parameters are considered to be highly correlated (ridge reg.).
#for(j in 1:7){ b[j]~dnorm(0,rho)}
#rho~dgamma(.0001,.0001)

#Model 3:  The parameters are considered random coefficients (robust reg.)

b[1:7]~dmnorm(mean1[],prec1[,])

#Hyperprior for mean1[]
mean1[1:7]~dmnorm(nm[1:7],Omega1[,])

#Hyperprior for the covariance matrix
prec1[1:7,1:7]~dwish(Omega2[,],7)
```

Appendix 8-Military and Network Efficiency Analysis

This appendix begins with the addition of another method, game theoretic DEA, which is used in the analysis of efficiency for between group comparison.

Game Theoretic Model

Another interesting and useful DEA formulation involves a competitive relationship among groups of DMU's. This approach may be quite useful in the analysis of the military health care system as one can eliminate the "within group" effect to evaluate the "between groups" effect. Specifically, the civilian (network) entities are often considered competitors to the military entities. Knowing the relative efficiency of civilian network components in comparison with only the military components might result in different referral patterns and vice versa.

The two-person ratio efficiency game may be written as follows (Rousseau & Semple, 1995):

$$\begin{array}{c|l} \text{Player} & \underset{(u,v)\in P}{Min}\ \underset{\lambda\in Q}{Max} \left\{ \dfrac{\sum_{j=1}^{n} (\lambda_j y_j)^T v}{\sum_{j=1}^{n} (\lambda_j x_j)^T u} \right\}, \\ 1 & \end{array}$$

$$\begin{array}{c|l} \text{Player} & \underset{\lambda\in Q}{Max}\ \underset{(u,v)\in P}{Min} \left\{ \dfrac{\sum_{j=1}^{n} (\lambda_j y_j)^T v}{\sum_{j=1}^{n} (\lambda_j x_j)^T u} \right\}, \\ 2 & \end{array}$$

$$P = \left\{ (u,v) : u \geq 0, v \geq 0, \frac{v^T y_0}{u^T x_0} = 1 \right\}$$

$$Q = \left\{ \lambda : \sum \lambda_j = 1, \lambda_j \geq 0 \right\}$$

The input and output vectors (x_j and y_j) are strictly positive in this formulation, and the vector x_j contains the m organizational inputs while the vector y_j contains the s organizational outputs. Organization "o" is the selected unit for evaluation (in this case, player 1).

In words, player 1 is a selected DMU from a collaborative team. Player 2 is the group of competitive organizations. Player 1 seeks to minimize the maximum payout to player 2 by selection of weights v and u, while player 2 seeks to maximize the minimum payout by selecting the appropriate λ's. If the payoff to player 2 is less than or equal to 1, then player 1 is considered ratio efficient. If the payoff to player 2 is greater than 1, then player 1 is ratio inefficient. This two-team game leads to a CCR game theoretic fractional program.

The fractional programming problem (Brockett, Chang, and Rousseau, 2004) follows. Consider two groups defined as G and G^C. In this formulation, each DMU belonging to one group is compared against the other group. For each DMU denoted k, the fractional programming problem formulation is:

$$\begin{aligned} &\underset{u,v}{Max} \quad \frac{y_k^T v}{x_k^T u} \qquad k \in G \\ &s.t. \\ &\qquad \frac{y_j^T v}{x_j^T u} \le 1 \qquad j \in G^C \\ &\qquad u, v \ge 0. \end{aligned} \tag{7}$$

Because the objective function is composed of elements of G and the constraint set is composed of elements of G^C, the total efficiency score may exceed one in this formulation. A DMU is ratio efficient if the optimal value of the objective function is greater than or equal to one. A DMU is ratio inefficient if the objective function is less than one. This analysis may prove to provide managerial relevance.

This appendix provides a comparison of the network versus the military facilities using multiple formulations (n=144). The purpose of the comparison is to glean information regarding network performance and Military Treatment Facility performance, which might be useful to managers in making patient referral and resource allocation decision. Within group comparisons (e.g., network to network) are useful in determining where patients might be referred for cost efficient and high quality care. Between group comparisons (e.g., network to MTF) are useful in evaluating the overall system performance of the Army Medical Department. It is useful to know how each MTF performs in comparison to the network components, understanding that the relative efficiency of all MTFs is likely to be lower than the network efficiency, as the readiness mission requires additional slack resources.

In this appendix, DEA efficiency estimates are derived using the traditional BCC-I model with the usual variables. Second, Game Theoretic DEA estimates are produced for comparison. Finally, individual network services are evaluated using BCC-I and a reduced set of inputs and outputs to identify potential referral pattern changes. The results of this chapter indicate that efficient components of one type (e.g., military facilities) are associated with inefficient components of the other type.

BCC Formulation and Results

The formulation for the BCC-I model follows.

$$Minimize\ \beta - \eta(es^-_D + es^+)$$
$$Y\lambda - s^+ = y_i$$
$$X_D\lambda + s^- = \beta x_i$$
$$X_{ND}\lambda + s^- = x_i$$
$$e\lambda = 1$$
$$\lambda \geq 0, \beta \geq 0$$

The input and output variables remain unchanged; however, there are limitations associated with separating the network from the military facilities, as discussed previously in Chapter 4. Descriptive statistics are provided in the next table and chart.

Table 55. Descriptive Statistics, Network and MTFs.

The descriptive statistics for all facilities follow. The median cost (budget) for civilian and military facility operations is \$51.95 million. The average technical efficiency score was .76. A breakout by military and civilian system component operations is provided in Figure 39.

	N						
	Valid	Missing	Mean	Median	Std. Deviation	Minimum	Maximum
Budget	144	0	96.9166	51.5950	115.8902	6.35	591.70
Enrollment	144	0	36.4320	28.1400	25.6440	.10	105.70
RWPs	144	0	22.5996	12.9100	28.3009	.96	127.37
RVUs	144	0	243.4554	154.1900	228.1495	11.54	957.51
Quality / Prev	144	0	87.4914	87.4950	3.5585	78.89	96.51
Access	144	0	2.9999	3.0300	.9824	.69	4.84
BCC Efficiency	144	0	.7634	.7729	.2081	.27	1.00

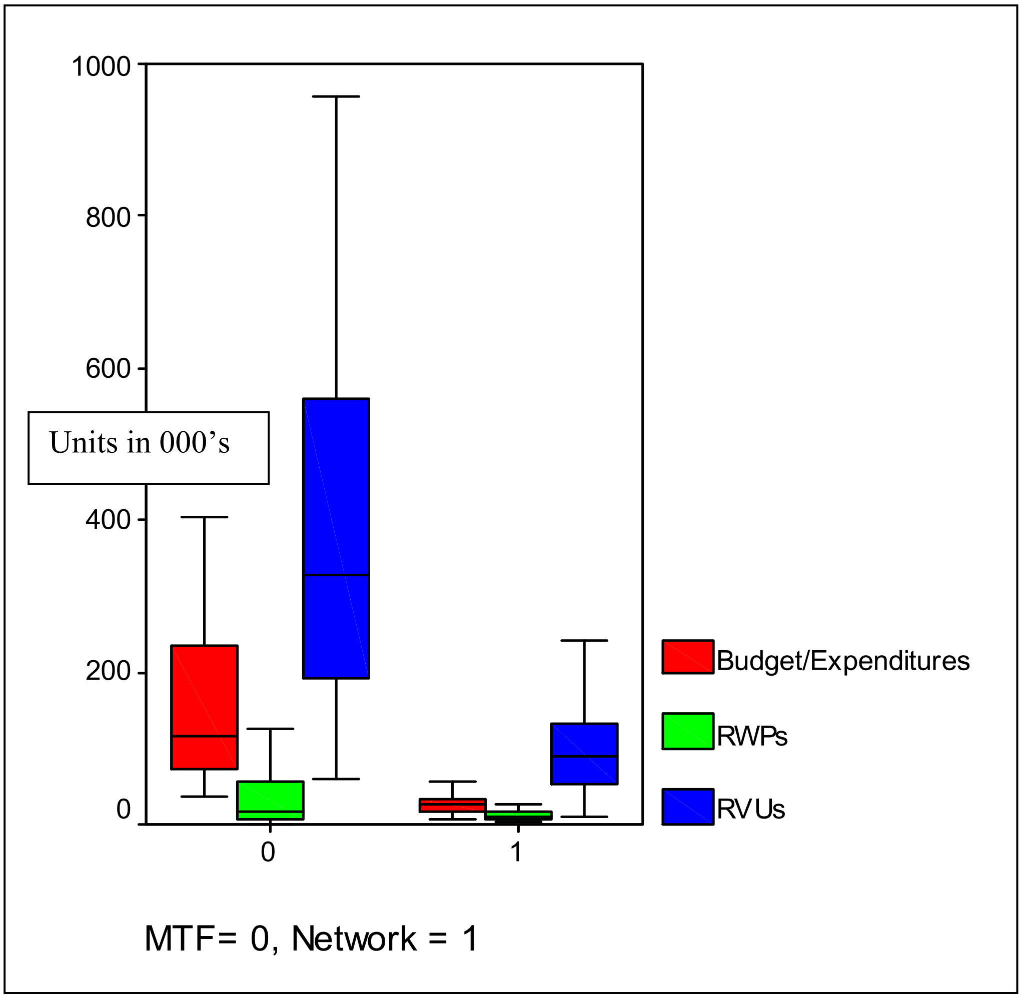


Figure 39. Boxplot of Network and MTF Variables.

The boxplot of the three variables below reveals that there is quite a bit more variation in the military facilities versus the civilian network components.

The results of the traditional BCC-I formulation for all three years are depicted in Figure 40. The most interesting phenomenon is that, for inefficient MTF components, there appears to be an efficient network, and for inefficient network components, there is an associated efficient MTF. In fact, the correlation is -.59 and statistically significant ($p<.05$). The ordering of military facility DMUs is consistent with the previous DEA analyses.

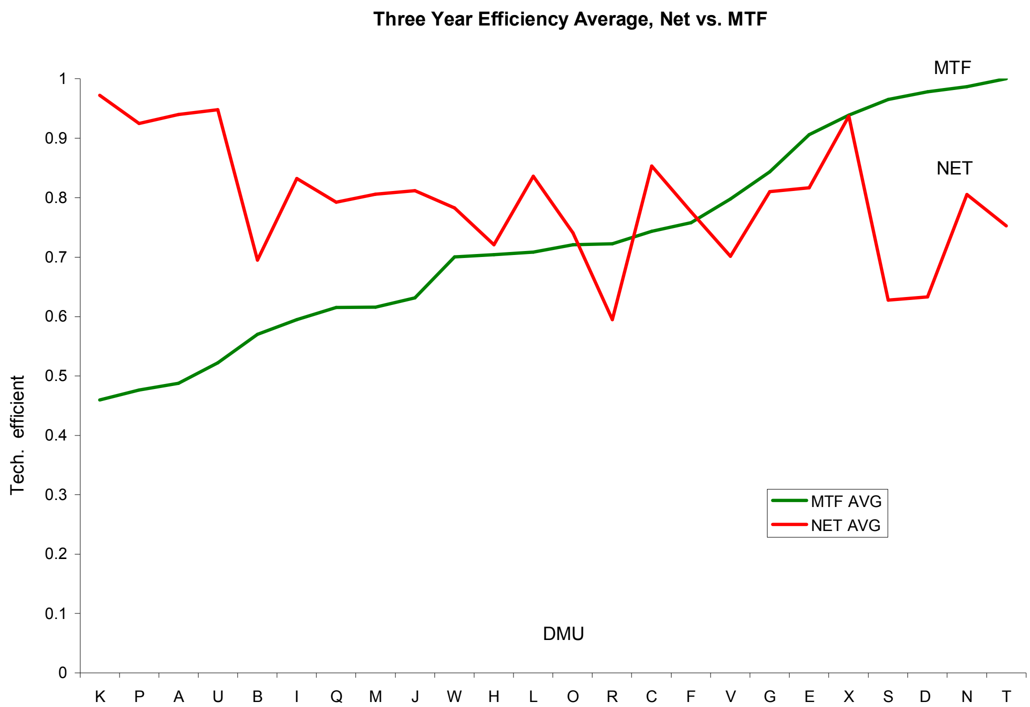


Figure 40. Plot of Average Technical Efficiency (Network and MTF).

The chart above illustrates an interesting phenomenon. It appears that inefficient network components are associated with efficient MTFs and vice versa.

The average efficiency scores suggest that negative correlation exists. Analysis of all hospitals for all three years as separate DMUs suggests that, while negative correlation exists, the networks generally perform better overall (see Figure 41).

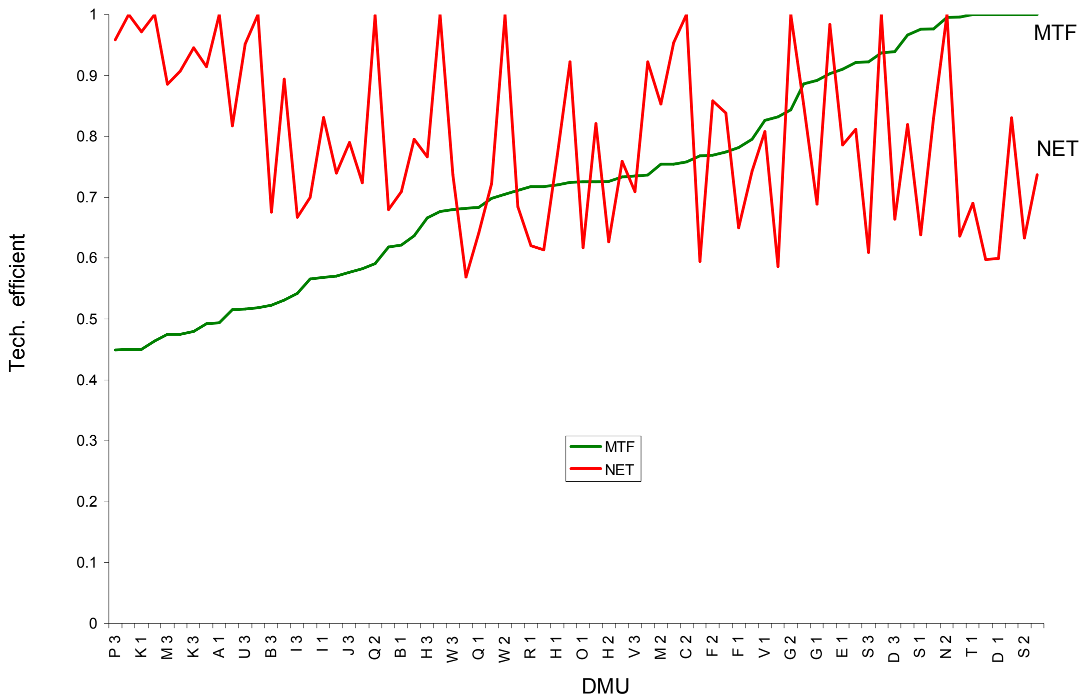


Figure 41. Plot of Observed Technical Efficiency, Network & MTFs.

The plot of all hospitals confirms the negative correlation between network and MTF performance; however, the network appears to perform (in general) better than the military facilities. The year of observation is indicated by the second number after the facility code (e.g., A1 is Facility A in 2001).

A comparison of the cost, quality, and access projections to the efficient frontier is congruent with previous analysis of military facilities. Table 55 provides a synopsis of the results.

Implications of the Analysis

Based solely on this analysis, it would seem useful for the military to try to recoup workload from the network around facilities R, V,G, E, S, D, N, and T based solely on Figures 37 and 38. The network performance around these facilities is less than that of the MTF performance. An analysis of these facilities reveals that six of these facilities are MEDCENs (so size improves efficiency relative to network performance), and the other two facilities are medium and large hositals. There is no geographic pattern; however, the MEDCEN factor is quite significant. The presence of a MEDCEN is associated with reduced network performance. The reasons for this reduced performance require some investigation, although it seems logical that very expensive cases (those outside the MEDCENs) capability are referred “downtown.” The costs of these cases could then make the networks cost inefficient relative to the MEDCENS.

Table 56. BCC-I Projections, 2001-2003, Network and MTFs.

The projections of cost, quality, and access (sorted by cost in 2003) follow. The network catchment area facilities are annotated with the prefix N and the suffix associated with the MTF that it supports.

	Cost 01	Quality 01	Access 01	Cost 02	Quality 02	Access 02	Cost 03	Quality 03	Access 03
P	-48.46%	1.36%	2.72%	-53.59%	0.30%	5.09%	-55.06%	33.33%	4.83%
M	-38.14%	0.00%	1.99%	-24.59%	0.00%	3.33%	-52.52%	0.00%	2.13%
A	-50.62%	0.00%	0.70%	-50.78%	0.00%	1.38%	-52.48%	33.33%	1.07%
K	-55.03%	0.00%	2.76%	-55.03%	0.00%	3.75%	-52.01%	33.33%	2.01%
U	-46.88%	0.00%	3.09%	-48.15%	33.33%	1.51%	-48.36%	33.33%	2.37%
B	-37.91%	3.77%	2.47%	-43.36%	36.75%	1.15%	-47.72%	35.74%	2.67%
I	-43.17%	4.22%	4.16%	-32.35%	0.00%	2.13%	-45.84%	38.76%	2.68%
NR	-37.95%	0.00%	3.81%	-40.52%	33.33%	1.23%	-43.10%	33.33%	1.58%
Q	-31.65%	0.00%	0.65%	-40.91%	4.07%	0.28%	-42.97%	0.00%	0.00%
J	-26.36%	0.00%	2.67%	-41.82%	2.56%	3.83%	-42.36%	311.45%	1.83%
NS	-36.14%	0.00%	0.00%	-36.76%	0.60%	1.59%	-39.05%	0.00%	1.44%
L	-24.56%	0.00%	0.17%	-26.68%	0.00%	0.00%	-36.37%	40.76%	0.60%
ND	-40.07%	0.00%	4.25%	-36.39%	35.34%	1.26%	-33.64%	36.95%	1.37%
H	-28.06%	8.58%	1.72%	-27.39%	8.89%	1.93%	-33.38%	7.38%	2.09%
NI	-16.89%	0.00%	0.00%	0.00%	0.00%	0.00%	-33.32%	0.00%	0.00%
NB	-29.06%	0.00%	3.33%	-30.04%	33.33%	0.59%	-32.41%	33.33%	0.97%
W	-28.24%	7.68%	5.07%	-29.51%	7.68%	2.30%	-32.07%	43.17%	3.32%
R	-28.25%	6.93%	2.63%	-23.19%	8.28%	2.73%	-31.82%	42.17%	0.49%
NO	-38.28%	0.00%	0.67%	-7.77%	0.00%	0.00%	-31.59%	0.00%	0.00%
C	-22.59%	10.69%	2.97%	-24.22%	47.39%	2.24%	-30.14%	337.35%	5.41%
NV	-19.23%	0.00%	0.23%	-41.42%	0.00%	0.25%	-29.09%	0.00%	0.00%
O	-27.52%	8.58%	1.00%	-27.54%	8.89%	2.42%	-28.88%	8.73%	0.74%
NC	-16.15%	0.00%	0.00%	0.00%	0.00%	0.00%	-27.82%	0.00%	0.00%
F	-21.90%	0.00%	2.70%	-23.10%	0.00%	1.72%	-27.46%	10.24%	1.30%

V	-17.38%	0.00%	0.98%	-16.81%	13.40%	4.45%	-26.51%	47.99%	5.23%
NT	-30.96%	0.00%	0.22%	-16.97%	17.12%	0.00%	-26.33%	0.00%	0.00%
NW	-38.64%	0.60%	1.11%	0.00%	0.00%	0.00%	-26.30%	0.00%	0.00%
NQ	-35.86%	0.00%	1.03%	-0.16%	0.00%	0.00%	-26.10%	0.00%	0.00%
NG	-31.13%	0.00%	0.60%	0.00%	0.00%	0.00%	-25.78%	0.00%	0.00%
NH	-23.26%	0.00%	0.30%	-37.32%	1.66%	0.30%	-23.34%	0.00%	0.00%
NJ	-7.80%	0.00%	0.00%	-27.64%	0.00%	0.00%	-21.01%	0.00%	0.00%
G	-10.87%	0.00%	0.00%	-15.67%	15.36%	2.97%	-20.45%	14.16%	0.66%
NL	-4.64%	0.00%	0.00%	-24.14%	0.00%	0.00%	-20.42%	0.00%	0.00%
NN	-40.23%	0.60%	3.54%	0.00%	0.00%	0.00%	-18.09%	0.00%	0.00%
NF	-35.01%	3.46%	0.58%	-14.13%	33.33%	0.00%	-17.88%	0.00%	0.00%
NE	-21.42%	7.68%	3.54%	-18.87%	44.18%	0.57%	-14.76%	46.39%	0.64%
NM	-32.07%	0.75%	3.03%	-14.67%	0.00%	0.42%	-11.45%	0.00%	0.00%
E	-8.94%	17.32%	2.76%	-7.90%	16.87%	1.82%	-11.35%	363.86%	1.92%
X	-2.34%	0.00%	3.02%	-6.28%	19.43%	3.80%	-9.71%	58.23%	2.12%
NA	0.00%	0.00%	0.00%	-8.61%	0.00%	0.00%	-9.30%	300.00%	0.14%
S	-2.41%	21.39%	0.14%	0.00%	0.00%	0.00%	-7.74%	47.62%	0.00%
D	0.00%	0.00%	0.00%	-0.37%	0.00%	0.61%	-6.11%	23.28%	0.00%
NK	-2.79%	0.00%	0.00%	0.00%	0.00%	0.00%	-5.52%	0.00%	0.00%
NU	-10.58%	0.00%	0.00%	0.00%	0.00%	0.00%	-4.82%	33.33%	0.38%
NP	-18.33%	0.00%	0.00%	0.00%	0.00%	0.00%	-4.19%	0.00%	0.00%
N	0.00%	0.00%	0.00%	-0.49%	22.75%	0.00%	-3.32%	61.45%	0.41%
NX	-17.03%	0.00%	0.00%	0.00%	0.00%	0.00%	-1.68%	0.00%	0.00%
T	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%

Game Theoretic DEA

The purpose of the game theoretic DEA analysis is to determine if there are any between group effects that might be of interest to managers. The last section in this appendix provides the GAMS model code for the GT DEA analysis. The game theoretic DEA formulation is a two-"person" (network vs. military) competitive game. Each component is thought to seek profit optimality at the expense of the other. (Expanded research would use an *n*-person game formulation.)

The results of the GT analysis are interpreted differently than the results of a typical BCC-I analysis. Specifically, those DMUs with efficiency scores greater than one are considered to be ratio efficient in comparison with the contrast group. One might expect that the network facilities would typically perform more efficiently than the military facilities, if only because of the built-in requirement for both the readiness and training missions in the latter. Even with cost adjustments for these missions, the secondary and tertiary effects are likely to be problematic.

The GT DEA analysis identifies the network components as more efficient than the military components. In fact, the military components are universally ratio inefficient, as would be expected due to the overhead of readiness and training requirements. For the network, however, it is interesting to note that catchment areas I, C, H, F, O, R, and X are inefficient for all three years of the analysis. The following table illustrates the results.

Table 57. Game Theoretic DEA, Networks vs. MTFs.

The GT DEA analysis illustrates that the MTFs are universally inefficient when compared to the network components, as would be expected. The differences in network efficiency might be used to change referral patterns and allocate FTEs. Facilities coded in red received scores less than one, while those in green received scores greater than one and are considered ratio efficient. (Note: a red "one" is a function of rounding; the facility scored just below "one.")

DMU	01 MTF	02 MTF	03 MTF	01 NET	02 NET	03 NET
M	0.460	0.446	0.419	1.118	1.086	1.069
K	0.444	0.444	0.459	1.356	1.301	1.180
P	0.480	0.456	0.447	1.030	1.037	1.014
I	0.469	0.478	0.450	0.967	0.962	0.954
U	0.482	0.464	0.455	1.227	1.715	1.328
A	0.494	0.490	0.469	1.364	1.268	1.266
Q	0.489	0.491	0.474	1.081	1.038	1.006
B	0.521	0.499	0.476	1.067	1.080	1.064
F	0.520	0.512	0.506	0.947	0.973	0.969
L	0.524	0.512	0.502	1.085	1.154	1.055
H	0.521	0.522	0.498	0.986	0.987	0.971
J	0.530	0.513	0.502	1.094	1.127	1.073
C	0.534	0.525	0.497	0.991	0.981	0.980
O	0.525	0.521	0.513	0.994	0.973	0.954
W	0.532	0.523	0.508	1.018	1.027	0.987
G	0.528	0.535	0.514	1.099	1.061	1.034
V	0.540	0.545	0.504	1.016	0.979	0.990
R	0.543	0.561	0.521	0.995	0.998	0.968
T	0.556	0.564	0.554	1.081	1.055	1.025
S	0.571	0.567	0.545	1.077	1.049	1.019
E	0.561	0.571	0.557	1.000	1.020	1.029
X	0.583	0.561	0.546	0.964	0.961	0.958
D	0.579	0.575	0.551	1.015	1.008	0.997
N	0.581	0.582	0.568	0.992	1.017	1.000

The implication of the GT DEA analysis is that it might be possible to adjust personnel referral patterns for underperforming network components. In fact, a separate efficiency analysis of 26 network services identifies specific subcomponents of each network.

Network Service Efficiency

A separate analysis of network services using the BCC-I formulation and a reduced set of variables serves to provide information regarding manning and referrals to AMEDD managers. Underperforming network areas may indicate the potential to recapture workload by reallocating personnel to meet the demand.

The variables used in the analysis include catchment area users by type (Active Duty, dependents, and retirees), government cost, and patient cost. In this case, the population variables are inputs in to the system, and the cost is the processed output. No discretionary variables were selected for use, so the formulation is similar to an advanced ratio analysis. A separate analysis was performed for each of the 25 services for all network DMUs for only the year 2003. The number of DMUs analyzed was 24.

A total of 25 clinics were analyzed and the results tabulated. The network surrounding Facility I appears to be the least efficient, with a total efficiency score of .60. The mean efficiency scores for the services suggests that nutritional services are the most inefficient with an average efficiency of .67 followed closely by hematology services with an average efficiency of .69. There may be some advantage to keeping as much of this workload in the fixed facility hospitals, depending on the cost structure of the specific MTF. The following table provides the mean efficiency by service, and the subsequent table provides the results for the analysis by catchment area and by service.

Table 58. Mean Efficiency Scores of Network Services, 2003.

The mean efficiency scores sorted from lowest to highest show that nutrition services and hematology services are generally inefficient

Nutrition	0.670	**Adv Reactions**	0.846	**Special Pediatrics**	0.887
Hematology	0.693	**Gastroenterology**	0.849	**Urology**	0.889
Rheumatology	0.757	**Neurosurgery**	0.851	**Allergy**	0.896
Opthalmology	0.770	**Gynecology**	0.857	**Thoracic Surgery**	0.898
Infectious Disease	0.791	**Orthopedics**	0.870	**Eyes, Ears, Nose, Throat**	0.904
Nephrology	0.806	**General Surgery**	0.879	**Endocrinology**	0.907
Cardiology	0.808	**Pulmonary / Respiratory**	0.883	**Dermatology**	0.934
Mental Health	0.830	**Obstetrics**	0.886	**Drugs**	0.950
Neurology	0.836				

Table 59. Mean Efficiency of Network Services by Catchment Area.

The table below provides an analysis of catchment area clinic efficiency for each of 26 clinics.

	Rucker	Wain	Carson	WRAMC	Gordon	Martin	Stewart	TAMC	Riley	Camp	Knox	Polk	Leon
Total	0.70	1.00	1.00	0.72	1.00	0.68	0.78	1.00	1.00	0.85	0.60	0.75	0.85
Gen Surg	1.00	0.96	1.00	0.87	1.00	0.85	0.60	1.00	0.74	1.00	0.61	0.81	0.71
EENT	0.82	1.00	0.85	0.55	1.00	0.74	1.00	1.00	1.00	0.70	0.90	0.55	1.00
Neur Surg	0.89	1.00	1.00	0.36	0.68	0.98	0.47	1.00	1.00	1.00	1.00	1.00	1.00
Ortho	1.00	0.61	1.00	0.56	1.00	0.64	0.99	1.00	1.00	0.69	0.75	0.75	0.74
Thoracic	0.80	1.00	0.78	1.00	0.94	1.00	0.95	1.00	0.44	1.00	0.86	0.84	0.68
Uro	1.00	1.00	0.83	0.83	0.94	0.54	0.65	1.00	1.00	0.72	0.80	0.77	1.00
OB	0.53	0.95	1.00	1.00	1.00	1.00	0.70	1.00	0.70	0.73	0.88	0.97	1.00
GYN	0.73	1.00	0.88	1.00	1.00	0.53	0.65	1.00	1.00	0.56	0.80	0.76	0.98
OPHTH	0.24	0.81	0.79	0.85	0.47	0.52	0.53	0.60	1.00	0.87	0.86	0.54	1.00
Mental	0.99	0.46	0.92	0.54	1.00	0.84	1.00	1.00	1.00	0.66	0.84	0.44	0.59
Drugs	0.80	1.00	1.00	0.89	1.00	0.88	1.00	1.00	1.00	0.99	0.85	0.91	1.00
Special Peds	0.96	0.65	1.00	0.64	0.82	0.60	0.86	1.00	1.00	0.75	0.79	1.00	1.00
Nephrology	1.00	1.00	0.79	0.55	0.77	1.00	1.00	1.00	0.61	1.00	0.74	0.59	0.83
Nutrition	0.22	1.00	1.00	0.60	0.59	0.18	0.24	0.34	0.32	0.19	0.83	0.48	0.49
Infectious	0.88	1.00	0.74	0.48	1.00	0.56	0.80	1.00	0.74	0.66	0.66	0.33	0.61
Neurology	1.00	1.00	0.98	0.35	1.00	0.76	1.00	1.00	0.81	1.00	0.77	0.81	0.70
Pulm / Resc	1.00	0.96	1.00	0.82	1.00	0.67	1.00	1.00	1.00	1.00	0.60	0.61	0.77
Rheum	0.58	0.83	1.00	0.44	1.00	0.52	0.92	1.00	1.00	0.80	0.78	0.65	0.42
Adv Reactions	0.81	1.00	1.00	0.67	0.90	0.72	0.99	0.94	0.75	1.00	0.77	1.00	0.98
Allergy	1.00	1.00	0.67	0.52	1.00	0.90	1.00	1.00	1.00	0.85	1.00	0.74	1.00
Cardiology	0.94	1.00	1.00	0.69	1.00	0.47	1.00	1.00	1.00	0.78	0.51	0.73	0.62
Derm	0.97	1.00	1.00	0.63	1.00	0.96	1.00	1.00	0.94	0.47	0.97	0.92	1.00
Endocrin	1.00	1.00	1.00	0.77	1.00	0.85	1.00	1.00	1.00	0.95	0.74	0.96	0.87
Gastro	0.84	1.00	1.00	0.47	1.00	0.82	0.95	1.00	0.78	0.85	0.60	0.77	0.76
Hema	0.38	1.00	0.41	1.00	1.00	0.51	0.55	1.00	1.00	0.57	0.79	1.00	0.33

	USMA	Bragg	Sill	Jack	Beau	BAMC	Hood	Eustis	Belv	MAMC	Irwin
Total	0.96	1.00	0.62	0.73	1.00	1.00	1.00	0.75	1.00	0.96	1.00
Gen Surg	0.93	1.00	0.85	1.00	0.53	0.86	1.00	1.00	1.00	0.78	1.00
EENT	0.87	1.00	0.98	0.99	1.00	1.00	1.00	0.77	0.98	1.00	1.00
Neur Surg	0.38	1.00	0.56	0.70	0.89	0.68	1.00	1.00	1.00	0.84	1.00
Ortho	1.00	1.00	0.75	0.99	1.00	0.98	1.00	0.92	0.64	0.87	1.00
Thoracic	1.00	1.00	0.50	1.00	0.78	1.00	1.00	1.00	1.00	1.00	1.00
Uro	0.92	1.00	1.00	0.93	0.66	1.00	1.00	1.00	1.00	0.74	1.00
OB	0.58	1.00	1.00	0.96	1.00	1.00	1.00	0.97	0.77	0.54	1.00
GYN	0.82	1.00	0.66	1.00	0.70	1.00	1.00	1.00	1.00	0.52	1.00
OPHTH	1.00	1.00	1.00	0.58	0.59	1.00	1.00	0.63	0.90	0.73	1.00
Mental	0.56	1.00	1.00	0.81	0.93	1.00	0.80	1.00	1.00	0.57	1.00
Drugs	0.75	1.00	0.89	0.89	1.00	1.00	1.00	0.95	1.00	1.00	1.00
Special Peds	0.85	1.00	1.00	0.88	1.00	0.81	1.00	0.73	0.96	1.00	1.00
Nephrology	1.00	1.00	0.63	1.00	0.47	0.69	0.73	0.82	1.00	0.34	0.80
Nutrition	1.00	1.00	1.00	0.98	0.93	0.45	1.00	0.26	1.00	1.00	1.00
Infectious	0.50	1.00	0.64	0.97	1.00	0.56	1.00	0.88	1.00	1.00	1.00
Neurology	0.51	1.00	0.52	1.00	0.80	0.90	1.00	1.00	0.76	0.41	1.00
Pulm / Resc	1.00	1.00	0.69	0.96	0.78	0.86	1.00	0.81	0.93	0.73	1.00
Rheum	0.56	1.00	0.76	0.50	0.50	0.88	1.00	0.84	0.53	0.67	1.00
Adv Reactions	0.42	1.00	0.91	1.00	0.61	0.75	1.00	1.00	0.73	0.36	1.00
Allergy	0.68	1.00	0.89	0.61	1.00	1.00	1.00	0.74	0.94	1.00	0.98
Cardiology	0.86	1.00	0.39	0.74	0.91	0.61	1.00	0.90	0.80	0.46	1.00
Derm	0.97	1.00	0.70	0.93	1.00	0.96	1.00	1.00	1.00	1.00	1.00
Endocrin	0.64	1.00	1.00	0.79	1.00	0.86	1.00	0.93	0.71	0.70	1.00
Gastro	0.76	1.00	0.77	1.00	0.89	0.76	1.00	0.75	0.83	0.76	1.00
Hema	0.61	1.00	0.38	0.62	0.58	0.34	1.00	0.80	0.43	0.34	1.00

Summary

The work in this appendix represents a start for future research. It is included to illustrate that comparisons between the network and the MTFs may provide insight into network referral problems. Evaluation of the efficiency scores provides important decision support information for determining personnel and resource allocation. For example, an inefficient network service may suggest that the AMEDD should increase its staffing and capability in that region to recapture costs. The GAMS formulation of the problem follows.

GAMS Coding

```
sets
                  d       mtf system components
                  d2      network system components
                  ds(d)  selected dmus for mtf
                  ds2(d2) selected dmus for network
                  j      inputs and outputs
```

```
                 i(j)  inputs only
                 o(j)  outputs only
*k(j) ndi


Parameter        data(d,j)   mtf     input  output
                 data2(d2,j) network input  output


                 ;

Variables        v(i)    input weights
                 u(o)    output weights
                 eff     efficiency CRS


positive variables u,v;

Equations
          primobj(d)         primal objective funtion
          constraint(d2)     primal constraint
          primobj2(d2)        primal objective function for the second set
          constraint2(d)     sum of inputs and ndi equal to one

          ;


*BCC-Input Oriented Model (VRS)
*Obj Fx:  E = Sum of weighted outputs / sum of weighted inputs for selected MTF DMU
primobj(ds)..        eff=e=sum(o,data(ds,o)*u(o))/(sum(i,data(ds,i)*v(i))+.0000001);
*Constraint:  Sum of weighted outputs is less than sum of weighted inputs for NETWORK
DMUs
constraint(d2)..  sum(o,data2(d2,o)*u(o))=l=(sum(i,data2(d2,i)*v(i)))  ;

*Obj Fx:  E = Sum of weighted outputs / sum of weighted inputs for selected MTF DMU
primobj2(ds2)..
eff=e=sum(o,data2(ds2,o)*u(o))/(sum(i,data2(ds2,i)*v(i))+.0000001);
*Constraint:  Sum of weighted outputs is less than sum of weighted inputs for NETWORK
DMUs
constraint2(d)..  sum(o,data(d,o)*u(o))=l=(sum(i,data(d,i)*v(i)))  ;


model            primvrs primal /primobj, constraint/
model            prim2vrs primal /primobj2, constraint2/
*The sets are currently defined to reflect cost as an input (budget flowing in)
*and those served coming out (Active Duty dependents, Retirees, etc.)
*This is the global model for all catchment areas around the major facilities.
.

sets             d       mtfs / H1*H72/
                 d2      networks / H73*H144/
                 j       inputs and outputs /Pop, Total, Work, Qual, Access/
                 i(j)    inputs            / Pop, Total   /
                 o(j)    outputs /  Work, Qual, Access /
*(k(j)    nondiscretionary /POP/  /


*Notes:  This is 2001-2003 data.

*MTFs follow

Table data(d,j)
          Pop         Total        Work        Qual              Access
H1        1.2         3.895        1.95        4                 4.28
….
```

```
H72       3.38        5.762        3.19        3                 4.14


*Networks follow
Table data2(d2,j)
           Pop         Total         Work        Qual     Access
H73        0.81        1.8486        1.49        4        4.24
….
H144       3.16        4.3268        2.45        4        4.39


$eolcom //

set ii(d) /H1*H72/
*Use this next line for CRS model
*convex.fx = 0;
;

parameter rep Summary Report;

loop (ii, ds(ii)=yes;


solve primvrs using nlp max eff;

rep(ii, 'Eff=') = primvrs.objval;
*rep(ii, 'Obj Fx=') = primvrs.objval;

ds(ii)=no);

set iii(d2) /H73*H144/
*Use this next line for CRS model
*convex.fx = 0;
;

loop (iii, ds2(iii)=yes;


solve prim2vrs using nlp max eff;

rep(iii, 'Eff=') = prim2vrs.objval;
*rep(ii, 'Obj Fx=') = primvrs.objval;

ds2(iii)=no);

display rep;
```

Appendix 9-Notation

$f(q)$	Probability Density Function (PDF)
$f(q,r)$	Joint Density Function of q and r
$F(q)$	cumulative distribution function (CDF)
$c = \beta_0 x_1^{\beta_1} x_2^{\beta_2} \ldots x_n^{\beta_n}$	Cobb-Douglas Production Function
H	Hessian matrix, matrix of 2d partial derivatives
Inf, Sup	Infimum, Supremum
$\lVert J \rVert$	Absolute value of the determinant of the Jacobian
$L(\beta \mid Y, X, \Sigma)$	Likelihood function
Log	Refers to the natural logarithm
$\underset{\beta,\lambda}{Min}$	Minimize with respect to β and λ
MVN	Multivariate Normal
N	Normal distribution
N^+	Half Normal Distribution
Φ	Standard Normal distribution
$\Re^m$	Real numbers (dimension m)
Σ	Covariance matrix
$U(0,1)$	Uniform distribution
$W_m(\Sigma, m)$	Wishart distribution
Y	Dependent variable vector
X	Independent variable matrix
∇	Gradient

Appendix 10-Acronyms

ACH	Army Community Hospital
AD	Active Duty
AMC	Army Medical Center
AMEDD	Army Medical Department
APV	Ambulatory Patient Visit
BCC	Banker, Charnes, and Cooper DEA model (see VRS)
CASS	Center for AMEDD Strategic Studies
CCR	Charnes, Cooper, Rhodes DEA model (see CRS)
CDF	Cumulative Distribution Function
CMAD	Case Mix Adjusted Discharges
CRS	Constant Returns to Scale DEA model
D	Discretionary input or output
DEA	Data Envelopment Analysis
DTIC	Defense Technical Information Center
DMU	Decision-Making Unit
FA	Factor Analysis
FTE	Full Time Equivalent
HMO	Health Maintenance Organization
IPA	Independent Practice Association
LOWESS	Locally Weighted Scatterplot Smoother
LS	Least Squares
MAC	Maximum Allowable Charge
MCMC	Markov Chain Monte Carlo
MEDCOM	Medical Command
M-H	Metropolis-Hastings Algorithm
MTF	Military Treatment Facility
MVN	Multivariate Normal distribution
ND	Non-discretionary input or output
OLS	Ordinary Least Squares
PCA	Principle Component Analysis
PDF	Probability Density Function
PPO	Preferred Provider Organization
RBRVS	Resource Based Relative Value System
RMC	Regional Medical Command
RVU	Relative Value Unit (Outpatient Weighted Workload)
RWP	Relative Weighted Product (Inpatient Weighted Workload)
SFA	Stochastic Frontier Analysis
VRS	Variable Returns to Scale DEA model
WHO	World Health Organization
WLS	Weighted Least Squares

Glossary

“AMEDD — Army Medical Department. All Army health care people in the world belong to the AMEDD. The Army surgeon general, a lieutenant general, sets policies for the entire AMEDD. "Dual-hatted" as U.S. Army Medical Command commander, he also commands most AMEDD commands, agencies and fixed hospitals.

Army Community Hospital — ACHs offer more complex, resource-intensive secondary care (e.g., inpatient care, surgery under general anesthesia) at major posts. Usually 50 to 150 beds. Also give primary care at outpatient clinics inside and outside the hospital, e.g., troop clinics and outlying clinics at small posts. A facility that offers all ACH services except inpatient care is called an Army Health Center.

Clinic — Outpatient facility offering primary care or simple specialty care, i.e., routine exams, tests and treatments. Supervised by a MEDCEN, MEDDAC or DENTAC. May be a stand-alone site (e.g., an Army Health Clinic) or part of a major health facility (family practice clinic, pediatric clinic, etc., within a hospital).

DENTAC — Dental Activity. Each DENTAC is the collocated health care partner of a MEDCEN or MEDDAC. DENTACs supervise local and outlying dental clinics and dental clinic commands. DENTACs at MEDCEN sites often are tertiary-care facilities with more specialists and special equipment. Smaller DENTACs refer complex cases to them.

Field medical units — Standardized Table of Organization and Equipment (TO&E) units, mobile and assigned to combat commands. They are not part of U.S. Army Medical Command, but MEDCOM helps with their clinical training and supervises their clinical work. Many MEDCOM people are "earmarked" to join these units in wartime.

Fixed treatment facilities — Table of Distribution and Allowances (TDA) units, with staffs and equipment tailored for specific sites/roles. In war, they remain in place and care for evacuated casualties. U.S. Army Medical Command units are TDA.

MEDCEN — Army Medical Center, offering tertiary care (sophisticated diagnosis/treatment of any ailment) as well as primary and secondary care. Each MEDCEN has a hospital plus other services (preventive medicine, blood bank, etc.). MEDCEN hospitals are larger than ACHs, have more sophisticated equipment and more specialized staffs, and offer wider arrays of specialty care. All MEDCENs offer graduate medical education (internships, residencies, etc.) for physicians.

MEDCOM — U.S. Army Medical Command, whose three-star commander is also Army surgeon general. Includes virtually all of the AMEDD except field units. Regional Medical Commands, or RMCs, are multistate command-and-control headquarters that allocate resources, oversee day-to-day management and foster readiness among MTFs in their areas. Regional Dental Commands and Regional Veterinary Commands perform similar command-and-control functions.

MEDDAC — Medical Department Activity, a medical command-and-control headquarters at a given post. A typical MEDDAC includes one ACH or clinic plus non-hospital elements (preventive medicine, blood bank, etc.), but the number of ACHs is not rigidly fixed. Also, not all ACHs belong to MEDDACs.

MTF — Medical Treatment Facility. Any Army medical care site, whether a clinic, hospital, medical center, etc.

RMC — Regional Medical Command. A multistate command-and-control headquarters that allocates resources, oversees day-to-day management and fosters readiness of MTFs in its region. Regional dental commands and regional veterinary commands perform similar command-and-control functions."

References

Aczel, Amir. (1993). Complete Business Statistics, 2d Edition. Irwin: Boston, MA.

Adult Technical Manual, Health Care Survey of Defense Beneficiaries (2003). http://www.tricare.osd.mil/survey/HCsurvey/.

Aigner, D.J.; Lovell, C.A.K., and Schmidt, P. (1977). Formulation and estimation of stochastic frontier production models. Journal of econometrics, 6: 21-37.

Army Medical Department Web. (2005.) A day in the Army Medical Department. [accessed on 15 February 2005]. Available at http://www.armymedicine.army.mil/about/intro/aday.html.

Ali, I. & Seiford, L (1993). Computational accuracy and infinitesimals in data envelopment analysis. Informs, 31: 290-297.

Arnold, V.L.; Bardhan, I.R.; Cooper, W.W.; & Kumbhakar, S.C. (1996). New uses of DEA and statistical regressions for efficiency evaluation and estimation-with an illustrative application to public secondary schools in Texas. Annals of Operations Research, 66:255-277.

Baltagi, B. (2001). Econometric analysis of panel data (2d edition). Wiley: West Sussex, England.

Bannick, R. R. & Ozcan, Y.A. (1995). Efficiency analysis of federally funded hospitals: Comparison of DoD and VA Hospitals using data envelopment analysis. Health Services Management Research, 8(2), 73-85.

Banker, R.; Charnes, A & Cooper, W.W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. Management Science, 30: 1078-1092.

Banker, R.; Conrad, R.; and Straus, R. (1986). A comparative application of data envelopment analysis and translog methods: an illustrative study of hospital production. Management Science, 32(1): pp. 30-44.

Bardhan, I., Cooper, W., & Kumbhakar, S. (1998). A simulation study of joint uses of data envelopment analysis and statistical regressions fro production function estimation and efficiency evaluation. Journal of Productivity Analysis, 9: 249-278.

Battese, G. & Coelli, T. (1992). Prediction of rim-plevel technical efficiencies with a generalized frontier production function and panel data. Journal of Econometrics, 38: pp. 387-399.

Beck, N.; Katz, J.; Alvarez, R.; Garrett, G.; Lange, P. (1993). Government partisanship, labor organization, and macroeconomic performance: a corrigendum . The American Political Science Review, 87:4, pp. 945-948.

Bertsimas, D. & Tsitsiklis, J. (1999). Introduction to linear optimization. Belmont, MA: Athena.

Box-Muller Transformation (2004). Wikipedia Online Encyclopedia. Retrieved on 1 July 2004 from http://en.wikipedia.org/wiki/Box-Muller_transformation.

Brockett, P.L.; Chang, R.E; Rousseau, J.J.; Semple, J.H.; & Yang, C. (2003). A comparison of HMO efficiencies as a function of provider autonomy. Pending publication.

Brockett, P.L.; Cooper, W.W.; Kumbhakar, S.C.; Kwinn, M.J.; McCathy, D. (2003). Alternative statistical regression studies of the effects of joint and service specific advertising on military recruitment. Pending publication.

Bureau of Labor Statistics (2004). Consumer price indexes. http://stats.bls.gov/cpi/

Casella, G. and Berger, R. (2000). Statistical Inference, 2d Edition. New York: Thomson Learning.

Casella, G. and George, E. (1992). Explaining the gibbs sampler. The American Statistican, 46:3, pp. 167-174.

Caves, D. and Tretheway, M. (1980). Flexible cost functions for multi-product firms. Review of Economics and Statistics, 62: 477-481.

Centers for Medicare and Medicaid (2004). On-line. Viewed January 2004. http://www.cms.hhs.gov/.

Charnes, A.; Cooper, W.; Dieck-Assad, M.; Golany, B.; & Wiggins, D. (1985) Efficiency analysis of medical care resources in the U.S. Army Health Service Command. Center for Cybernetic Studies, Research Report CCS 516. Contract number NSF-SES84-08134. University of Texas at Austin. Distributed by the Defense Technical Information Service (ADA 159742).

Charnes, A; Cooper, W.; & Rhodes, E. (1978) Measuring the efficiency of decision making units. European Journal of Operational Research, 2. pp. 429-444.

Charnes, A.; Cooper, W.W.; and Sueyoshi, T. (1988). A goal programming / constrained regression review of the bell system breakup. Management Science, 34: 1-26.

Chib, S., and Greenberg, E. (1994). Bayes inference for regression models with ARMA(p,q) errors. Journal of Econometrics, 64, 183-206.

Chib, Siddhartha, Greenberg, Edward (1995). Understanding the Metropolis-Hastings algorithm. The American Statistician, 49:4, 327-335.

Cleveland, W. (1979). Robust locally weighted regression and smoothing scatterplots. Journal of the American Statistical Association, 74: 829-836.

Coelli, T. (1995). Estimators and hypothesis tests for a stochastic frontier function: a Monte Carlo analysis. Journal of Productivity Analysis. 6:4, pp. 247-268.

Coppola, N. (2003). Correlates of military medical treatment facility (MTF) performance: measuring technical efficiency with the structural adaptation to regain fit (SARFIT) model and data envelopment analysis (DEA). Unpublished dissertation. Virginia Commonwealth University.

Cook, D. R. & Weisberg, S. (1999). Applied Regression Including Computing and Graphics. New York: Wiley.

Cooper, W. L.; Seiford, M.,;& Tone, K. (1999). Data envelopment analysis. Boston: Kluwer Academic Publishers.

Cooper, W.W.; Seiford, L.M & Zhu, J. (2004). Handbook on data envelopment analysis. Boston, MA: Kluwer.

Cornwell, C.; Schmidt, P.; & Sickles, R. (1990). Production frontiers with cross-sectional and time-series variation in efficiency levels. Journal of Econometrics, 46:1/2, pp. 185-200.

Coventry, J.; Gromadzki, R; Hutchinson, K.;Kiernan, M.; Rogers, S.; Smith, G. & Spivey, P. (1995). MHSS Workload Primer. Unpublished manuscript. Available at http://www.TRICARE.osd.mil/tma/hpae/primword.html#2.2.5.

Creel, M. (2004). Graduate Econometrics Lecture Notes. Accessed on 17 January 2005. http://pareto.uab.es/mcreel/Econometrics/.

Cremieux, P. & Ouellette, P. (2001). Omitted variable bias and hospital costs. Journal of Health Economics. 20: 271-282.

Damien, P.; Wakefield, J.; & Walker, S. (1999). Gibbs sampling for Bayesian non-conjugate and hierarchical models by using auxiliary variables. Journal of the Royal Statistical Society 61:2, pp. 331-344.

Dor, A. (1994). Non-minimum cost functions and the stochastic frontier: On applications to health care providers. Journal of Health Economics, 13(3), 329-334.

Farrell, M.J. (1957). The measurement of productive efficiency. Journal of Statistics and Social Series 120, 253-281.

Ferrier, G.D. & Lovell, C.A.K. (1990). Measuring cost efficiency in banking-econometric and linear programming evidence. Journal of Econometrics, 6:229-245.

Foster, G. and Gupta, M. (1990) Manufacturing overhead cost driver analysis. Journal of Accounting and Economics, 12: 309-337.

Fox, J. (2002). Time series regression and generalized least squares. Appendix to An R and S-PLUS Companion to Applied Regression. Sage Publications: California.

Gerdtham, U.; Loethgren, M.; Tambour, M; & Rehnberg, C. (1999). Internal markets and health care efficiency: a multiple-output stochastic frontier analysis. Health Economics 8: 151–164.

Goldberg, M. (2001). Cost and Efficiency Measurement for Military Hospitals. Contract study.

Greene, W. (2003). Econometric analysis, 5th Edition. Prentice Hall, Upper Saddle River: NJ.

Greene, W. (2003). Distinguishing Between Heterogeneity and Inefficiency: Stochastic Frontier Analysis of the World Health Organization's Panel Data on National Health Care Systems. (2003). Paper presented to Current Developments in Productivity and Efficiency Measurement, University of Georgia, October 25-26, 2002 and the Second Hellenic Workshop on Productivity and Efficiency Measurement in Patras Greece, May 30, 2003. Available at http://pages.stern.nyu.edu/~wgreene/heterogeneityandinefficiency.pdf

Hastings, W. (1970). Monte Carlo sampling methods using Markov Chains and their applications. Biometrica, 57: 97-109.

Health Care Survey of DoD Beneficiaries (HCSDB) Overview (2004). http://www.tricare.osd.mil/survey/HCsurvey/.

Hillier, F. & Lieberman G. (2001). Introduction to Operations Research, 7th Edition. Tata McGraw Hill: New Delhi.

Hofler, R.A. and S.T. Folland, (1991). Technical and allocative inefficiencies of United States hospitals under a stochastic frontier approach, for presentation at the Midwest Economics Association. Fifty-fifth Annual Meeting (St. Louis, MO).

Introduction to Monte Carlo Methods (1995). Computational Science Education Project, Department of Energy. Retrieved on 1 July 2004 from http://csep.hpcc.nectec.or.th/CSEP/MC/MC.HTM.

Jeffreys, H. (1961). Theory of Probability. Oxford U. Press.

Jondrow, J.; Materov, I.; Lovell, K.; & Schmidt, P. (1982). On the estimation of technical inefficiency in the stochastic frontier production function model. Journal of Econometrics, 19 (2/3), pp. 233-238.

Kumbhakar, S. & Lovell, C (2000). Stochastic Frontier Analysis. Cambridge University Press.

Kumbhakar, S. C. (1990). Production frontiers, panel data, and time-varying technical inefficiency. Journal of Econometrics, 46:1/2, pp. 201-212.

Lee, L. and Schmidt, P. (1993). A production frontier model with flexible temporal variation in technical inefficiency. In Fried, Lovell And Schmidt, eds. The Measurement of Productive Efficiency: Techniques and Applications. New York: Oxford.

Magnusson, J. (1993.) Efficiency Measurement and the Operationalization of Hospital Production. Health Services Research, 31 (1): 21-37.

Mantha, V. (2004). Factor analysis and multidimensional scaling. Unpublished manuscript. mantha@isip.msstate.edu

MEPRS Basic Course (2004). Unpublished manuscript. Available at: ampo.amedd.army.mil/meprs/pdfs/MEPRS Basic Tng Stu Man.pdf

Meeusen, W. & van den Broeck, J. (1977). Efficiency estimation from Cobb-Douglas production functions with composed error. International Economic Review, 18:2, 435-44.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equations of State Calculations by Fast Computing Machines. Journal of Chemical Physics, 21, 1087-1092.

Metzger, W. (2002). Statistical methods in data analysis. Katholieke Universiteit Nijmegan. Experimental High Energy Physics Group. On-line text.

Mihara, T. (1990). Measuring the efficient utilization of personnel at Navy military treatment facilities. A Thesis for A Master of Science in Operations Research, Naval Post Graduate School.

Ozcan, Y.A. & Bannick, R.R. (1994). Trends in Department of Defense hospital efficiency. Journal of Medical Systems, 18(2), 69-83.

Prais, S.; & Winsten, C. (1954). Trend estimation and serial correlation. Cowles Commission Discussion Paper Number 383, Chicago.

Reinhard, S.; Lovell, K. & Thijssen, G. (1997). Environmental efficiency with multiple environmentally detrimental variables; estimated with SFA and DEA. Wageningen Economic Papers, issue 07-97.

Roberts, G.O., Gelman, A., and Gilks, W. R. (1994). Weak Convergence and Optimal Scaling of Random Walk Metropolis Algorithms. Technical Report. University of Cambridge.

Ross, Sheldon (2001). A First Course in Probability 6th Edition. New York: Prentice Hall.

Rousseau, J.J. & Semple, J. (1995). Two-person ratio efficiency games. Management Science, 41:3, pp. 435-441.

Spiegelhalter D J, Best N G, Carlin B P and van der Linde A (2002). Bayesian Measures of Model Complexity and Fit (with discussion). Journal of the Royal Statistical Society, 64, 583-640.

Stoloff, P.H.; Lurie, P.M.; Goldberg, L; Kimko, D; & Almendarez, M. (2004). Evaluation of the TRICARE program, 2002 report to Congress. Retrieved July, 2004. Available at www.TRICARE.osd.mil.

TMA Statistical Reports (2004). Retrieved 2 January 2005. Available at http://199.211.83.250/Reports/.

Thurstone, L.L. (1947). Multiple-factor analysis. Chicago: University of Chicago Press.

Tierney, L. (1994). Markov Chains for Exploring Posterior Distributions. Annals of Statistics, 22, 1701-1762

West, Timothy (2000). AMEDD Cost Efficiency by facility. Contract study.

Yalcin, A. & Amemiya, Y. (2001). Nonlinear factor analysis as a statistical method. Statistical Science, 16(3): pp. 275-294.

Zuckerman, S; Hadley, J.; Iezzoni, L. (1994). Measuring hospital efficiency with frontier cost functions. Journal of Health Economics, 9: 1-21.

VITA

Lawrence Van Fulton is a senior military officer, medical evacuation pilot, student, and native Texan. Born in 1966 to Bill and Dell Fulton, he attended schools in San Antonio, Texas. Upon graduation, he attended Texas State University and graduated with a Bachelor of Science degree. After entering active duty military service, he served as an officer and aviator. During his 20-years of service, he has completed graduate degrees in Statistics from the University of Texas at Austin, Health Care Administration from Baylor University, Military Arts and Sciences from the Command and General Staff College, Human Resource Management from Troy State University. Currently, he has two published papers on file in the Defense Technical Information Center and is pursuing his PhD through the University of Texas at Austin.

Permanent Address: 112 Valona Drive, Cibolo, Texas 78108

This dissertation was typed by the author.